The London School of Economics and Political Science

Fertility, Education and Social Mobility in 20th Century Colombia

Juliana Jaramillo-Echeverri

A thesis submitted to the Department of Economic History of the London School of Economics and Political Science for the degree of Doctor of Philosophy, London, February 2023.

Abstract

The main objective of this thesis is to contribute to our understanding of the fertility transition, the unequal opportunities in access to education and the long-term persistence of social status in twentieth-century Colombia.

The thesis has two parts. The first part presents a detailed empirical perspective of the rapid Colombian fertility decline including results at the national, sub-national and individual-level. Using full individual-level census data, the results show that before the fertility transition fertility norms differed in the country because of the different historical legacies. But by 1964 fertility started falling simultaneously across all regions and fertility halved almost everywhere in only 25 years; the only exception was the indigenous communities that did not undergo a fertility decline during this period. This sudden decline suggests that a technological or cultural shock could explain the decline. However, the findings confirm that knowledge and access to family planning did not bring about a faster fertility decline. I provide new stylised facts for the long-run relationship between female education and fertility. At the individual level, the relationship between education and fertility holds strongly. However, at the national and sub-national level, fertility decline cannot be explained by the direct effects of education as fertility fell continuously in all educational groups since 1965.

The second part of this thesis provides new evidence on the unequal access to education and presents a long-term view of social mobility using surname-based methods. The results show that contemporary patterns of social exclusion in education are rooted in the colonial past. However, the long-term effects of the colonial past have been uneven. While the historical elite reinvented itself, and some have been slowly losing their elite status, indigenous and Afro-Colombians have been trapped with scarce possibilities to move upward. The results also reveal low social mobility and a segregated marriage market with high levels of homogamy at the top and the bottom of the social ladder.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of 69,063 words, excluding references.

Statement of co-authored work

I confirm that Chapter 5 was jointly co-authored with Professor Andrés Álvarez and I contributed 50% of this work.

Acknowledgements

Gracias quiero dar al divino
Laberinto de los efectos y de las causas
Por la diversidad de las criaturas
Que forman este singular universo,
Por la razón, que no cesará de soñar
Con un plano del laberinto,
Por el rostro de Elena y la perseverancia de Ulises,
Por el amor, que nos deja ver a los otros
Como los ve la divinidad,

Otro poema de los dones
Jorge Luis Borges

This thesis concludes a long research journey that started under the supervision of Neil Cummins and Eric Schneider. Since I met them, Neil and Eric inspired a big deal of this thesis with their own research interests. They have patiently accompanied me in my research with an open and curious mind to learn about Colombia, and were both sharp and kind when reading and commenting on the many iterations of this manuscript. Our meetings and conversations were fundamental for my PhD thesis and my future life as an economic historian. I feel lucky to have met and learned from such scholars, and I want to thank them for their rigour and generosity.

The journey of a PhD always begins years before its official start. I came to love economic history thanks to Juan Esteban Velez and I have shared this passion with my beloved coauthor Julieth Andrea Parra and lifelong colleagues like Javier Mejia and Adriano Franco, who also read and commented on this thesis. My love for economic history further developed under the guidance of Maria Teresa Ramirez and Adolfo Meisel. They taught me how to rigorously combine both economic logic

and knowledge of history to conduct research. Salomon Kalmanovitz gave me the opportunity to teach economic history for the first time. During my time at the Central Bank of Colombia, I shared with Daniel Osorio my love for real analysis, London, the LSE and economic history questions. Other friends supported me during those times. I owe special thanks to Carlos Brando, Manuela Barrera, Alberto Boada, Jaime Bonet, Xavier Duran, Edwin Lopez, David Tobon, Marcela Ocampo, and Melba Velasquez.

I have a deep debt of gratitude to the matchless academic environment of the Department of Economic History at the London School of Economics. During my time in the department, my research benefited from insightful discussions with Gerben Bakker, Gregory Clark, Sara Horrell, Chris Minns, Mary Morgan, Albrecht Ritschl, Max Schulze and Patrick Wallis. Alejandra Irigoin was a dear friend and a constructive critic, our discussions stimulated my love and curiosity for Latin American issues.

My friends and colleagues during these years improved this research and made my life light and fun. Aurelius Noble and Mario Cuenda read several parts of this thesis and provided editorial as well as moral support during these years. My office mates Alka, Alex, Andrea, David, Felix, Greta, Hillary, Ivan, Jenny, Juan Jose, Mattia, Manik, Mauricio, Sam, Sijie, Tan, and Victor made research interesting, challenging and fun, even during those long days in the lockdown. My Oxford friends, Andres Guiot and Luis Gabriel Galan have shared with me the complete trip.

I was able to apply to and fund a PhD thanks to the financial support of the Central Bank of Colombia, which funded my first 4 years. I benefited from the financial support of the Department of Economic History with the Radwan Travel and Discovery Fund. I am also grateful to Neil Cummins, Jordan Claridge, Leigh Gardner, Karolina Hutkova, and Eric Schneider for the opportunity to gain experience

and money while working with them. The support from Helena, Jeannie, Tracy and Lorraine made all the administrative tasks of the PhD life, effortless and pleasant. I will take in my heart forever Loraine's magnificent cakes.

This thesis benefited greatly from the interaction with other scholars in countless seminars, workshops and conferences. I owe special gratitude to the attendees of the conferences and seminars organised by Young Demographers, the Economic History Association, the British Society for Population Studies, the Population Association of America, the Economic History Society, the departments of Economic History at LSE, Lund University, and Queen's University Belfast during 2020 and 2021. From these conferences I obtained useful and stimulating feedback, while we were all in our rooms, tired of being connected to the world only through our screens.

I also thank the organisers and participants of the workshop on Social mobility at Bocconi University, the seminars series of the Latin American Center at the University of Cambridge, the European Historical Economics Society Annual Conference, the European Society for Historical Demography Conference, the Uppsala Economic History Graduate Seminar, the CID seminar at Universidad Nacional de Colombia, and the World Economic History Congress. Guido Alfani, Hoyt Bleakley, Naim Bro, Diego Castañeda, Pablo Fernández Cebrián, James Fenske, Maria Jose Fuentes, Eilidh Garrett, Jonas Helgertz, Grant Miller, Ramon Ramon Muñoz, Federica Querin, and Stephanie Thiehoff took the time to engage with my work and gave me invaluable feedback. Among the best conferences, I want to thank Per Engzell and Giacomo Gabbuti for the co-organisation of a wonderful online seminar in 2020 from which I learned a lot. We also had a beautiful session in Paris in the summer of 2022. I enjoyed those days greatly.

I also want to thank my examiners, Alice Reid and Santiago Perez, for agreeing to read and evaluate this thesis. They read carefully my manuscript and

we had a challenging and stimulating discussion. Their comments, questions and suggestions were key for the final version of this thesis. My friends and family gave me unconditional love and support and they always made me feel at home. I learnt during these years that all flourishing is mutual. I want to thank especially Andres, Ari, Laura, Marco and Vero for their love, without which, this thesis would not exist. Ari Velez revised this thesis with an editorial eagle-eye. Andrés Álvarez has been a wonderful coauthor and accomplice during these years. Laura Luna provided the best meme materials. If life is but a dream, Marco Martinez is the best dreaming partner in the universe. Living in London with Veronica Salazar was beautiful. Ethan, Olivia, Mia and Duna were born during these years. Their presence brightens my days.

Terminar un doctorado se siente como un triunfo, aunque puedo reconocer la importancia del azar y la buena fortuna. Comparto este triunfo con mis tías y con mis papás. De ellos aprendí la perseverancia, la pasión, y la compasión, que han sido cruciales en mi vida. Giulio es una fuente inagotable de alegría y amor. Su interés, preguntas, críticas y, en especial su paciencia, mejoraron esta tesis, y mi día a día, de manera indescriptible.

Contents

List of Figures

List of Tables

1

Introduction

The world experienced radical demographic and economic transformations in the 20^{th} century. Fertility rates dropped nearly in all countries, while economic growth steadily increased the living standards of almost all humans. Although these are evident changes for almost everyone, the opportunity to enjoy these changes has not been equally distributed across countries and individuals. Low and middle-income countries have struggled with high fertility rates and low social mobility longer than Western European economies.[1] Although it is certain that fertility decline and equal-

[1]The relationship between inequality and fertility has been a matter of extended debate since the early 19^{th} century, when fertility started declining in Europe. For example, De La Croix and

ity of opportunity are linked to improvements in socioeconomic development, the causes behind these transformations are still a matter of debate in the literature.

This thesis focuses on the demographic and economic transformations of Colombia during the 20th century. Looking at the case of Colombia is interesting as the country has been characterised in the literature by its large inequalities at the regional and the individual level (Galvis & Meisel, 2014; Narayan et al., 2018; OECD, 2018). In fact, in 2018 the country was considered one of the most unequal countries in the world. Colombia seems to have achieved modernisation and witnessed considerable social change during the 20th century in terms of reduction of poverty, women's rights, gains in life expectancy and education. Also, the country experienced one of the fastest fertility declines in the world. However, these transformations did not translate into more equal opportunities for Colombians. Strikingly, studies estimate that in Colombia it would take at least 300 years for the offspring of low-income families to reach the mean income (OECD, 2018; Ramirez-Zuluaga, 2016).

The main objective of the thesis is to fill gaps in our knowledge of the fertility transition, the unequal opportunities in access to education and the long-term persistence of social status in Colombia. The thesis has two parts consisting of five chapters total. The first part presents a detailed empirical perspective of the rapid Colombian fertility decline including results at the national, sub-national and individual-level. The second part of this thesis provides new evidence on the long-term persistence of social status in Colombia and the unequal opportunities for access to high-quality education, looking in particular at elite groups and Afro-Colombians and indigenous.[2]

Colombia is a multi-ethnic country, and it is home to the second-largest

Doepke (2003) argue that fertility differentials can account for the relationship between inequality and growth. See also Malthus (1803) and Clark (2007).

[2]Throughout the thesis, I refer to Afro-Colombians and indigenous people as ethnic groups when I mention both categories together.

black population in South America and to 115 native indigenous groups. Colombia's wide extensive land area, as well as the Andean mountain ranges have presented considerable barriers to the integration of the regions of the country and between the country and the outside world. It is the third most populous country in Latin America and the fourth-largest economy in the region by gross domestic product (GDP) and, in terms of GDP per capita it appears just above the average. Most of its economic growth started after the 1870s while urbanisation, schooling and industrialisation took off during the first part of the 20$^{\text{th}}$ century. In terms of the determinants of the economic development of the country, the literature has highlighted the role of the export sector (Ocampo, 2004), industrialisation (Kalmanovitz, 1983) and the long persistence of colonial institutions (García-Jimeno, 2005). Less attention has been placed on understanding other indicators of development and the unequal distribution of economic growth.

Several books tracing the economic history of Colombia leave out two important topics: the demographic transition –in particular the rapid fertility decline of the country–, and the evolution of inequality and social mobility.[3] Instead, almost all these books include an analysis of foreign trade, the agricultural sector –in particular the coffee sector–, the emergence of industry and the banking system, the fiscal and monetary policy history, the construction of railways and highways, the emergence of mass education and the slow but steady economic growth. The omission of chapters studying the demographic transition and inter-family inequality from most textbooks is due to relatively scarce data that has limited both fields of study.[4] Fortunately,

[3]A quick view into the indices of the most prominent books on Colombian economic history reveals the absence of these topics. See for example: Palacios and Safford (2012), Bushnell (1996), Melo (2017), Tirado Mejía (1998) and Robinson and Urrutia (2007). An exception is a compilation by Meisel and Ramírez (2010) which includes one chapter on demography since the colonial times, and Kalmanovitz (2015) who includes one chapter on inequality and income distribution at the regional level. There are various books that examine these topics, but they have not been fully integrated into the economic history literature. See for example Flórez-Nieto (2000) for an analysis of demographic change, and Urrutia and Berry (1975) for a history of inequality in Colombia.

[4]For example, the lack of official birth registers from 1980 to 1998 or the absence of longitudinal

recent developments in methods and the digitisation of full-count historical censuses make it possible to study demographic and economic processes at an unprecedented scale and detail.

The first part of the thesis studies one of the fastest fertility declines in the world. Although the mechanisms behind fertility decline are widely unknown, the historical fertility transition has been fertile ground for research. For the case of Western Europe and the US, several authors have focused on the relationship between fertility decline and the emergence of economic growth (Galor & Weil, 2000; Voigtländer & Voth, 2006; Clark, 2007; Dennison & Ogilvie, 2014; Guinnane, 2011). This line of research argues that changes in the economic environment brought about changes in fertility as parents responded to economic incentives (Galor & Weil, 2000; G. Becker, 1960). In particular, some attention has been given to changes in the returns of education or in the opportunity cost of fertility (Bleakley & Lange, 2009; Tan, 2019; S. Becker, Cinnirella & Woessmann, 2010; Fernihough, 2017). Others stress the importance of cultural factors and the diffusion of new social norms across countries. The findings of the European Fertility Project reveal that fertility declined more or less at the same time, in countries with different economic conditions (Coale & Watkins, 1986). Following this idea of cultural transmission of new social norms, recent research has argued that fertility started declining first in French and then in English-speaking countries (Blanc & Wacziarg, 2020; Beach & Hanlon, 2019).

For the Latin America case, more attention has been placed on the role of contraception and family planning programs as the timing of the decline coincides with the introduction and popularisation of modern contraceptive methods like the pill in the mid-1960s (Cleland, 1985; Bongaarts, Mauldin & Phillips, 1990; Miller, 2010; Miller & Babiarz, 2016). Finally, other research has focused on the direct effects of women's education on fertility decline as female education has been historically

data that contains a measure of social status.

negatively associated with fertility (Castro Martin, 1995; Caldwell, 1980; Batyra, 2016).

In Colombia, the main demographic transformations in the 20th century started after the 1920s. Mortality declined around 1930 and fertility started declining in the 1960s. By 1990 the Total Fertility Rate was 3 children per woman and today is just below the replacement rate (1.8) (Flórez-Nieto, 2000; López Toro, 1968).[5] Mejía, Ramírez-Giraldo and Tamayo-Castaño (2008) investigate if the demographic transition was spurred by economic forces, as predicted by Galor and Weil (2000). They conclude that the structural change of the 1950s was translated into an increase in the demand for more educated workers and afterwards into fewer but more educated children. However, conclusive explanations about the rapid fertility decline of the country are still missing in the literature, as well as a deeper analysis of the regional differences in the fertility transition.[6]

As the literature has argued, a sub-national analysis is fundamental for understanding the fertility transition, given that looking only at the national fertility decline can hide relevant patterns in which fertility relates to other variables (Guinnane, 2011; Garrett, Reid, Schürer & Szreter, 2001; Dribe & Scalone, 2014). Sub-national data allow us to uncover regional differences within a country while individual-level data allow us to understand better the relationship between economic status and fertility. This is important in a country like Colombia which has been shaped by its geography and spatial disintegration. In fact, several studies have shown that regional inequality has been been a continuous phenomenon, as gaps in income and living standards have persisted since the 18th century (Bonet & Meisel-Roca, 2001;

[5]For the demographic transition in the Latin American context see Guzman, Singh, Rodriguez and Pantelides (1996).

[6]There has been noticeable more research on the mortality side of the demographic transformation. See for example Acosta-Ordoñez (2014) on infant mortality on a regional approach, Jaramillo-Echeverri, Meisel-Roca and Ramírez-Giraldo (2019) on explanations behind the decline in mortality rates and the increase in life expectancy, and Miller and Urdinola (2010) on the relationship between coffee exports and mortality.

Fergusson, Molina, Robinson & Vargas, 2017).

Although the regional perspective has attracted the attention of researchers on several topics, the analysis of the fertility transition has been mostly limited to a national perspective. Probably the main exception was the work by Gutiérrez de Pineda (1968), which inspired a big part of this thesis. In her ethnographic work, she visited several areas of Colombia and catalogued them based on their family organisation. She describes 5 types of families in Colombia: *Coastal*, *Andean*, *Antioqueña*, *Santandereana* and *Indigena*.[7] She also acknowledges that fertility patterns were changing in the country and that migration had the potential to spur this change. From an economic history perspective, the main limitation of Gutiérrez de Pineda (1968) was that she provided little empirical evidence on the fertility decline and the causes behind it.

The first part of the thesis presents new analyses of fertility, before and during the transition, at the sub-national level, which was missing from the literature. Additionally, it explores potential mechanisms behind the rapid fertility transition. My empirical analysis in the first three chapters is based on individual-level data from the complete census of 1973 and 1993. I also collect individual-level data from a 10% sample of the censuses of 1985 and 2005 from the Integrated Public Use Microdata Series – IPUMS International (Minnesota Population Center, 2019).[8] As for the methodology, I adopt the Own Child Method to estimate fertility rates from 1958 to 1990. To address potential differences across regions in the country, Chapter 2 tests

[7]Following Colombia's current administrative division, the *Coastal* cluster is composed by: Guajira, Magdalena, Atlantico, Cesa, Bolivar, Sucre, Cordoba, Choco, and parts of Antioquia, Valle del Cauca, Cauca and Nariño. The *Andean* cluster includes Boyaca, Cundinamarca, Huila and parts of Santander, Cauca and Nariño. The *Antioqueña* region includes Calda, Risaralda, Quindio and parts of Antioquia, Valle and Tolima. The *Santandereana* region is defined by the departments of Santander and Norte de Santander. Gutiérrez de Pineda (1968) excludes several departments from the analysis. For example, the indigenous communities are only those living in the Andean region, while the departments of Amazonas and Orinoquia are missing. See Table 2.1 and Fig. 2.2.

[8]I want to thank Steven Ruggles and all of IPUMS members in the University of Minnesota for such an amazing platform.

for the presence of geographical clusters before and during the fertility transition using Local Indicators of Spatial Association – LISA. The results show that before the fertility transition, variations in fertility patterns were linked to differences in historical legacies that derive from differences in geographical endowments. Then, I provide new estimations of fertility rates at the departmental level and for indigenous women. The findings reveal that fertility started falling simultaneously across all regions around 1964 and fertility halved almost everywhere in only 25 years; the only exception was the indigenous communities that did not undergo a fertility decline during this period.

The simultaneity of the decline across all regions indicates that the decline happened independently of economic development, which was certainly unequal across the country. Therefore, the next two chapters delve into other likely causes behind this rapid and widespread fertility transition. Chapter 3 investigates the effect of a national radio campaign promoting family planning clinics in the late 1960s. In this chapter, I exploit exogenous variation in radio signal strength at the municipal level, and I also employ a difference-in-differences strategy to compare fertility rates before and after the start of the radio campaign. The results show that there was no effect of the radio campaign on fertility rates across the country. When the analysis is restricted to municipalities that had a family planning clinic nearby, the results confirm that access to family planning had a limited effect. This chapter corroborates that knowledge of and access to family planning and contraception alone does not bring about a faster fertility decline. The findings also suggest that the potential effects of media and communication technologies on fertility rely more on the promotion of new social roles than on the dissemination of information about family planning.

In Chapter 4 I provide new stylised facts on the relationship between female education and fertility at the national, sub-national and individual-level. I focus on

the implementation of educational reforms in Colombia between 1930 and the late 1940s (in particular during the Governments of Olaya Herrera and Lopez Pumarejo), during the so-called liberal period. These reforms sought to promote mass education in the country placing important attention on female education. However, the total number of students enrolled in primary and secondary education did not start growing until after the 1950s. The findings indicate that the relationship between women's education and fertility is not necessarily monotonic. At the individual level, the relationship between education and fertility holds strongly. However, at the national and sub-national levels, fertility decline cannot be explained by the direct effects of education as fertility fell continuously in all educational groups since 1965. This suggests that forces other than education were driving the decline. Although fertility declined continuously, the results also show that the fertility gap between uneducated women and women with access to secondary education changed little. Fertility declined for women with secondary education, from 6 children to 2, whereas for women with no education fertility dropped from 8 children in 1960 to 4 in 1990.

The second part of this thesis investigates the long-term persistence of social status. There is a large debate in the literature regarding the persistence of social status and historical social mobility.[9] Recent research in the US and European countries has benefited from the new methods of automated record linkage and the digitisation of full count historical censuses (Abramitzky, Mill & Pérez, 2020; Abramitzky, Boustan, Eriksson, Feigenbaum & Pérez, 2021). The most recent estimations for the US show that historical social mobility was low with an estimated father-son association of around 0.6 to 0.8 between 1840 and 1910 (Ward, 2021). For the English case, Long and Ferrie (2013) estimate a father-son association of around 0.3 between 1851 and 1901, although some have suggested that this estimation could be biased

[9]For a review of the literature see Black and Devereux (2011), Solon (2018), Ward (2021) and Santavirta and Stuhler (2020). For the debate see Solon (2018); Clark, Cummins, Hao and Diaz-Vidal (2015); Guell, Mora and Telmer (2013) and Torche (2014).

downwards due to measurement error (Clark et al., 2015; Ward, 2021). Pérez (2019) studies occupational mobility in the late 19th century in Argentina by computing matrices of father-son occupational transitions for Argentina, the US, the UK and Norway. He finds that occupational mobility levels in Argentina were above those of Britain and Norway and comparable to those of white people in the US.

Unfortunately, for most Latin American countries, including Colombia, there is no possibility to use censuses to link generations in the past because in most full digitised censuses the name of the person is unknown. The scarcity of longitudinal data has therefore limited our knowledge of social mobility patterns in the long run, although historians suggest that the rigid social structure of the Spanish colonial regime is reflected in contemporary Colombia (Prados De La Escosura, 2005; Williamson, 2010; Bértola & Williamson, 2017; Dobado Gonzalez & Garcia Montero, 2010).

To overcome these data limitations, I use surname-based estimations. Surnames are an abundant source of data, as these almost always pass from one generation to another and carry relevant characteristics that are strongly related to social status. Family names have been used to identify the linguistic (Novotnỳ & Cheshire, 2012), ethnic (Mateos, Longley & O'Sullivan, 2011), genetic (King & Jobling, 2009), and socioeconomic (Bro & Mendoza, 2021) composition of populations, and also to measure long-term effects of colonial institutions in Peru (Carpio & Guerrero, 2021). In particular, surnames are especially relevant for the literature on historical social mobility (Guell, Rodríguez Mora & Telmer, 2007; Clark et al., 2015; Santavirta & Stuhler, 2020).

Andrés Álvarez and I have been collecting several lists of individuals and family surnames from indigenous pre–Hispanic groups, 17th century Spaniard colonial officers (*encomenderos*), enslaved Africans, members of privileged families with access

to higher education in the 17th and late 18th century, slaves owners in 1851, and members of different social and business elites of the late 19th and the beginning of the 20th centuries. The data construction is still a work in progress and our hope is to collect more data for Colombia, but also other Latin American countries. We think that this new data will enable future research to track persistence and social mobility in the presence of fragmented information.

In Colombia, the demographic and economic transformations of the 20th century affected the distribution of economic growth and access to opportunities, both at the regional and the individual level. Even though there is limited research on the long-term patterns of inequality and social mobility, the pioneering work of Londoño (1995) estimates Gini coefficients of income for the period 1938 to 1985.[10] Londoño's estimations show that from 1938 to 1960 inequality grew continuously from a Gini of 0.45 to a Gini of 0.56 and then it declined to 0.47 in 1985. According to Londoño (1995), the reverse in inequality after the 1960s was due to the spread of mass education and higher returns to education. Nonetheless, the reduction in inequality was transient. Inequality started rising rapidly once again in the 90s, achieving the highest levels in the century (0.6).

As of today, Colombia is regarded as one of the most unequal countries in the world, with an estimated Gini of 0.52. Although the literature argues that these patterns have roots in the colonial past, there is limited empirical evidence to support such claims. Therefore, how persistent are the elites of the past is still an open question. Similarly, most of the research on social mobility studies cohorts that were born after the second half of the 20th century, which limits our understanding of the long-term process of transmission of social status.[11]

[10]The Gini coefficient measures income concentration. It is presented between 0 and 1, where 0 is perfect equality and 1 is total inequality. In practice today most countries move around 0.3 to 0.6.

[11]See for example: Angulo, Azevedo, Gaviria and Páez (2012), Ramirez-Zuluaga (2016), Moyano and Galvis (2014), Galvis and Meisel (2014), Dahan and Gaviria (1999).

This thesis shows that the distribution and representation of surnames in the educational system confirm contemporary educational exclusion. The findings corroborate that these patterns are rooted in the past and have persisted over several generations. The results also show that the persistence of status varies across population groups. In particular, the results show that while the historical elite reinvented itself, and some have been slowly losing their elite status, ethnic groups have been trapped with scarce possibilities to move upward. As in the US, ethnic groups in Colombia show the worst social mobility outcomes (Ward, 2021).

In Chapter 5, Andrés Álvarez and I study the persistence of social groups from different historical periods and ethnic backgrounds linked with a clear position on the social ladder in the past. Social status is measured using surnames and their relative representation in access to low and high-quality education. By evaluating if a surname is over-represented or under-represented in high or low-quality education, the chapter presents empirical evidence on social segregation in the educational system today and concludes that this segregation reproduces patterns of social exclusion that are rooted in the colonial past. We show that the long-term effects of the colonial institutions were uneven across elite and ethnic groups. The chapter also studies the marriage market in Colombia for elite and disadvantaged ethnic groups. The results show there is virtually no matching between the elites and ethnic groups which reveals a highly segregated marriage market with high levels of homogamy at the top and at the bottom of the social ladder. This segregated marriage market matches the segregation of the educational system.

Chapter 6 analyses intergenerational mobility using graduates from Los Andes University from 1949 to 2018 and the social groups defined in Chapter 5. I estimate lower bounds of the intergenerational elasticity of social status. These new estimations allow us to understand long-term patterns of social mobility in Colombia

by linking the present to the past. The findings of this chapter confirm the low-mobility patterns found in other studies for Colombia with an estimated intergenerational coefficient of 0.52 for the whole sample. When I look at each historical group separately the results uncover different patterns of social mobility with coefficients that range from 0.45 to 0.6. Long-surviving elite families show lower social mobility in comparison to elite families that belong exclusively to the pre-industrial period or new elite families.

Chapter 7 discusses the broader lessons of the thesis, their policy implications and sheds light on new avenues of research.

El problema de la familia colombiana es bastante complejo, porque nosotros tenemos culturas muy diferenciadas, abrigadas bajo unas mismas fronteras.

Tres Culturas Familiares

Estanislao Zuleta

Part I

The rapid transformation: fertility decline

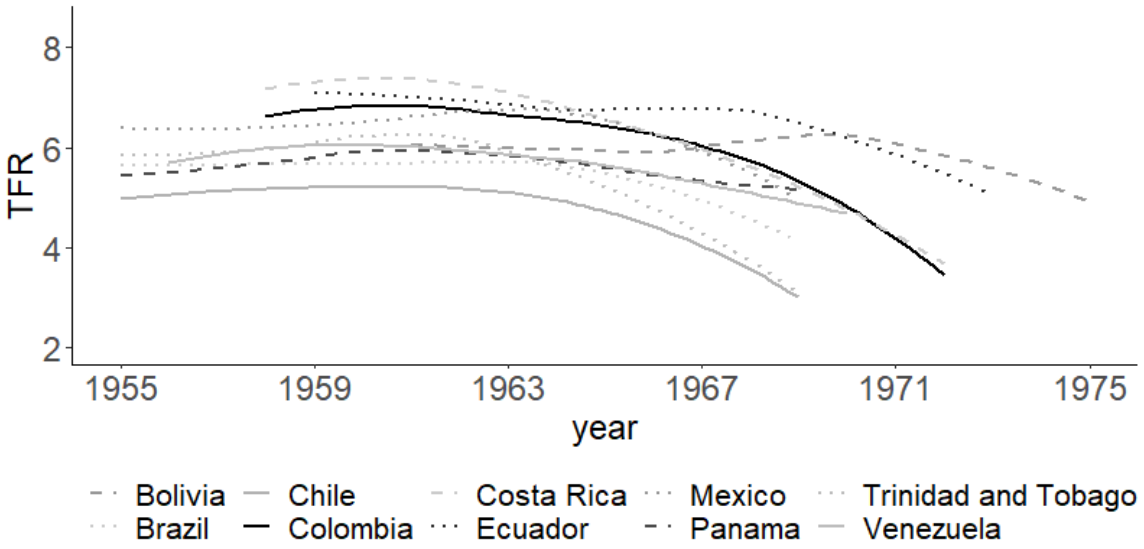
2

Fertility decline: new evidence from a regional approach

2.1 Introduction

Colombia experienced one of the fastest declines in fertility in the world: children per woman fell from 7 in 1960 to 3 in 1985 (Flórez-Nieto, 2000; Flórez, 1992).[1] Other Latin American and Caribbean countries -like Costa Rica and Brazil- also experienced rapid changes during the same period (1964-1993) as shown in Fig. 2.1. On the whole, the Latin American fertility decline was fast, unpredictable, and present in dissimilar countries. Most countries experienced fertility decline after the 1960s, except for Guatemala and Panama, which showed a high and stable fertility level, and Argentina and Uruguay, which had low and stable fertility since the late 19th century (Guzman et al., 1996).

Figure 2.1: Total Fertility Rate, selected Latin American countries



Note: Calculations are made with the Own Child Method. Women between 15 to 78 years old and children between 0 to 14 years old. See Section 2.6 for a description of the methodology. Sources: Authors' calculations based on Census samples from IPUMS-International (Minnesota Population Center, 2019), Mortality tables from Human Life-Table Database from 1970, and data from Colombia from the complete census 1973, National Agency of Statistics (DANE). The years of the censuses are 1970 for Brazil, Chile, Mexico and Venezuela, 1971 for Trinidad y Tobago, 1973 for Costa Rica, 1974 for Ecuador and 1976 for Bolivia. I released the implementation of the OCM to construct these fertility rates under an open source license (`https://github.com/jje90/ASFR_TFR`).

[1] During the historical fertility transition in Western Europe or the US, a similar drop –from more than 7 children per woman to fewer than 3– took more than 80 years (Guinnane, 2011)

What explains the fertility transition? Looking at the European fertility transition, several authors have focused on the relationship between fertility decline and the emergence of economic growth (Galor & Weil, 2000; Voigtländer & Voth, 2006; Clark, 2007; Dennison & Ogilvie, 2014; Guinnane, 2011). Cross-country evidence confirms that there are strong correlations between variables such as income and fertility (the highest fertility rates are found in the poorest countries) or fertility and education (women with more years of education have fewer children).

However, regionally disaggregated data shows that in developing countries, these demographic processes do not fully reflect the improvements in socioeconomic conditions. Specifically, convergence in fertility rates is not always linked with convergence in economic development between regions, which suggests a weak relation between demographic outcomes and economic performance. Studies that came after the World Fertility Survey (WFS) in 1973 cast "considerable doubt on the causal primacy of economic forces, either at the societal or family level, in bringing about fertility change in the contemporary third world" (Cleland,1985, p.5). Despite a generalised decline in fertility, there were important socioeconomic differences among (and within) countries. Overall, the survey shows that the fertility transition around the world did not exhibit a unique pattern. For the European case, the European Fertility Project (EFP) found that there were no regularities in socioeconomic variables and fertility among European provinces, as fertility declined more or less at the same time in countries with different economic conditions (Coale & Watkins, 1986).

In England, fertility decline was widespread and homogeneous with few regional differences, while in Germany these variations in fertility were more noticeable (Garrett et al., 2001; Knodel, 2015; Wilson & Woods, 1991). In France, the regional differences in fertility during the fertility decline are explained by secularisation although regional fertility differentials disappeared after the 1850s (Blanc, 2020;

Spolaore & Wacziarg, 2022; Daudin, Franck & Rapoport, 2019). For the US case, the literature has found north and south regional differences during the fertility transition, although some argue that this pattern was replaced by an East-West differential after the 1970s (O'Connell, 1981; Pandit & Bagchi-Sen, 1993). These findings suggest that to fully understand how fertility patterns and socioeconomic development are related, we have to analyse these relationships at a more disaggregated level.

Using the complete registers of the censuses of 1973 and 1993, I present a detailed empirical perspective of Colombia's rapid decline in fertility focusing my attention on the regional perspective. In Colombia, regional differences have been strongly related to geography and up until the late 19$^{\text{th}}$ century regions remained isolated which created an "archipelago economy" (Nieto Arteta, 1999). For example, Andean regions have different geographical endowments in comparison to plains, riverside or sea-coastal regions. It has been argued that differences in resource endowments led to different colonisation processes after 1500 which resulted in differences in institutions (Sokoloff & Engerman, 2000). It is also well-established that the lack of transport infrastructure maintained the regions isolated well into the 20$^{\text{th}}$ century. The Amazon region, despite being flat, with several navigable rivers and great natural wealth, had a late and unfinished colonisation process and is virtually disconnected from the economic and political centre of the country or the coastal region.

To understand the association between geography and fertility patterns before and during the fertility transition, I identify spatial patterns by testing for the presence of geographical clusters using Local Indicators of Spatial Association – LISA, for women with completed fertility before the fertility transition (born between 1910-1920) and during the fertility transition (born between 1930-1940).[2] Then I present new estimations of fertility from 1958 to 1990 at the national and the departmental

[2]Women with completed fertility are women that are unlikely to have another pregnancy. In this chapter I select women that were 53 to 63 years old by the time of the census.

level, computed using the Own Child Method.[3]

Across the thesis I focus on the fertility rates of all women instead of focusing on marital fertility, as has been done for the European case. The main reason is that in Colombia, as well as in other Latin American countries, levels of cohabitation and children out-of-wedlock are high and increased during the 20th century. In fact, between 1973 and 2005, the percentage of 25–29-year-old cohabiting women increased from 20% to 66% (Esteve, Saavedra, López-Colás, López-Gay & Lesthaeghe, 2016). Even more, from 1973 to 2005 the percentage of children born inside marriage declined across the country from around 50% in 1973 to less than 10%. On the contrary, children born in cohabiting couples increased from 30% in 1973 to more than 60% in 2005. Therefore, in this context, it makes sense to include all women in the analysis, despite their marital status.

The results confirm that, before the fertility transition, variations in fertility patterns were linked to differences in historical legacies that derive from differences in geographical endowments. However, fertility started falling simultaneously across all regions after 1964, and fertility declined as fast in places where women had traditionally low fertility as in places where women had higher fertility. Although the regions did not achieve convergence in fertility rates by 1983, the TFR halved almost everywhere in only 25 years. The simultaneity of the decline across all regions confirms that demographic processes are not always linked to improvements in socioeconomic conditions and suggests that other factors, such as the diffusion of social norms or technological changes, could have played a more prominent role in the fertility decline.

My results contribute to the debate on the origins of the fertility decline by looking at a country with wide varying socioeconomic characteristics that underwent a fast fertility decline across different groups of women. When studying the emergence

[3]Departments are similar to states in US or counties in England. There are formed by municipalities which are the smallest subdivision of the country.

of the fertility transition in Europe, the European Fertility Project (EFP) argued that the spread of new moral and cultural norms was responsible for the decline in fertility, as it occurred simultaneously in different socioeconomic contexts (Coale & Watkins, 1986).[4] Furthermore, recent research has provided empirical evidence on the effect of cultural factors on the fertility transition by tracing how social norms emerge and were spread in French and English-speaking countries (Blanc & Wacziarg, 2020; Beach & Hanlon, 2019; Spolaore & Wacziarg, 2022; González-Bailón & Murphy, 2013; Daudin et al., 2019). My results show that differences in fertility patterns before the fertility transition are explained by geographical and historical factors.

The Unified Growth Theory (UGT), based on Galor and Weil (2000) and Galor (2005), concludes that the transition from a high fertility regime to a low fertility regime is sparked by a change in parents' preferences, moving from quantity towards quality.[5] The model suggests that during the Industrial Revolution, as technological innovation took place and returns to human capital increased, parent's investment in children's education (quality) grew and fertility (quantity) declined.[6] The model implies that an increase in income and higher rates of return in quality explains the decision of parents to prefer fewer, better-educated children. However, this model fails to explain the fertility decline that occurred in France prior to the French Industrial Revolution. Additionally, the empirical literature on the existence of a quality-quantity tradeoff finds mixed results, especially regarding the studies of the historical fertility transition (Guinnane, 2011; Bleakley & Lange, 2009; Tan, 2019; Baez, 2008). The simultaneity and speed of the fertility decline in Colombia suggest that it cannot be explained by the expansion of education or the structural transformation of the economy. Instead, sudden technological or cultural shocks might

[4]See Guinnane, Okun and Trussell (1994) for a methodological critique of the EFP.

[5]The model also assumes that this is accompanied by a transition from a Malthusian world to a modern growth regime.

[6]Similar to the seminal work of G. Becker and Lewis (1973) and G. Becker (1993) that argues that parents maximise a utility function of quantity and expenditure on children (quality).

explain the decline.

2.2 Colombia: a country of regions

Colombia is located in the north of South America and currently, it covers an area of 1,141,748 square kilometres. It is bordered by the Caribbean Sea to the north and the Pacific Ocean to the west, and it shares borders with Venezuela, Brazil, Ecuador, Peru and Panama. The country lies across the equator between latitudes 12°N and 4°S with constant temperatures throughout the year that vary mainly according to the altitude. Given the substantial variation in elevation, shown in Fig. 2.3a, from sea level up to 5,730 m a wide set of climatic conditions are found across the territory (Bell, 1921). For example, Bogotá the capital of the country is located above 2.600 m of elevation and has an annual average temperature of 13°C. In contrast, Cartagena located over the Caribbean Coast has an average temperature of 29°C.

The country could be divided into 5 macroregions: Caribbean, Pacific, Andean, Orinoquia, and Amazonia based on their proximity and geographical characteristics. The departments that form each of these macroregions share similar historical and cultural characteristics. Topographic factors have been tied to the location of the population since pre-colonial times and were determinant in the foundation of settlements during the colonisation process. Also, these geographical features such as Los Andes, kept the regions isolated before the end of the 19$^{\text{th}}$ century when investments in roads and railroads started to grow.[7] Table 2.1 describes these macroregions, and Fig. 2.2 shows the map of Colombia placing these regions.

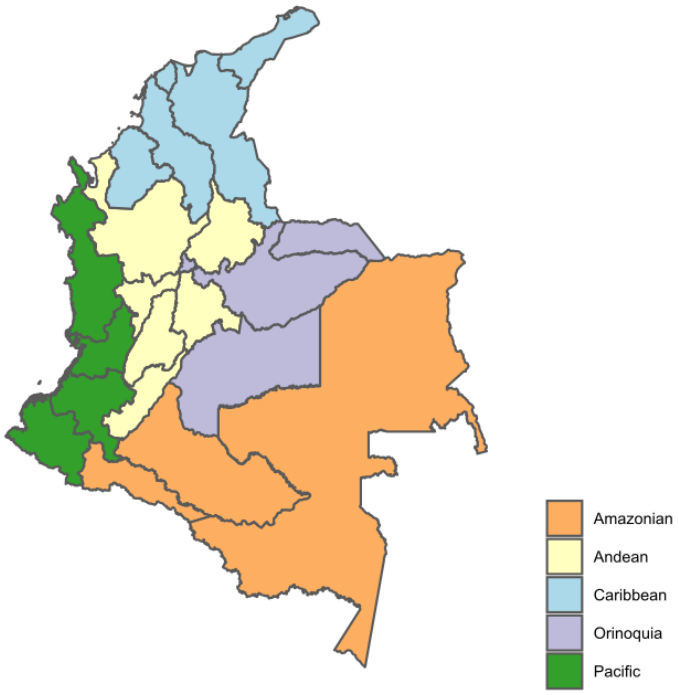
[7]Although scholars have argued that such infrastructure was primarily designed to connect towns and villages within the same region or to allow the exports of coffee and not to integrate the different regions (Gallup, Gaviria & Lora, 2003; Meisel-Roca, Ramírez-Giraldo & Jaramillo-Echeverri, 2016).

Table 2.1: Colombian macroregions

Region	**Location**	**Historical legacy and ethinc composition**	**Departments**
Caribbean	Along the Caribbean Coast	Important cities and ports during the colonial period. Historically defined by weak property rights which translates in high levels of poverty and informality. *Mestizo* (mix of Indigenous and Hispanic), black and traces of indigenous population.	Atlántico, Bolívar, Cesar, Córdoba, La Guajira, Magdalena, Sucre
Pacific	Along the Pacific Coast	Sparsely populated during the colonial period. The northern part was historically the mining district, while in the southern part haciendas were established. After the Independence free enslaved people settled in the northern part of region.	Chocó, Valle del Cauca, Cauca, Nariño
Andean	Andean mountain range	Inhabited by indigenous that later served to established the encomienda. The region was the centre of the colonial bureaucracy. The western Andean region remained isolated from the rest of the colony. Expansion of the territory during the 19^{th} century. Mainly Spanish origin.	Antioquia, Boyacá, Caldas, Cundinamarca, Huila, N. de Santander, Quindío, Risaralda, Santander, Tolima
Orinoquia	Located in the Orinoco River watershed	Sparsely populated as it includes 4 protected areas with presence of indigenous. Recent expansion of the frontier and discovery of oil reserves resulted in high land concentration. Indigenous people, Hispanic, and a mix of both (*mestizos*)	Arauca, Casanare, Meta, Vichada
Amazonian,	Amazon rainforest	Covers about 40% of the Colombian territory and is the least populated area in the country. Predominantly indigenous population	Amazonas, Caquetá, Guainía, Guaviare, Putumayo, Vaupés

Notes: If the division is made only in geographical terms some departments would appear in two macroregions. For example, the higher municipalities in Valle del Cauca are geographically located over the mountain Andean range, and share similar cultural traits.

Figure 2.2: Colombia divided into 5 macroregions



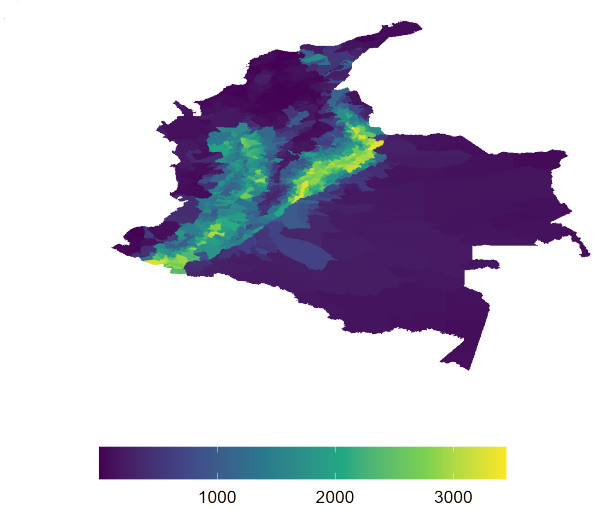
Notes: These macroregions are defined in Table 2.1 grouping departments by their geographical location and historical context. The boundaries are those of 1973.

The history of the country has been shaped by its geographical characteristics and spatial fragmentation (Palacios & Safford, 2012; Gallup et al., 2003; Bonet & Meisel-Roca, 2001). Pre-colonial indigenous settlements were found across the territory, but the Andean and the Caribbean regions had the highest density of indigenous population at the arrival of the Spaniards (Jaramillo Uribe, 1964; Melo, 1996). The first settlements during the Spanish conquest were in the high altitude regions which provided an escape from the heat and access to fertile lands and indigenous labour-force (Colmenares, 2007). Later, the expansion of the frontier moved from the interior to the lowlands. These colonisation patterns can be seen in Fig. 2.3b. In general, municipalities over the mountain Andean range, above 2,400 m of elevation, were the first colonial towns and became important bureaucratic centres. The port city of Cartagena and a few small towns located near the Magdalena river in the

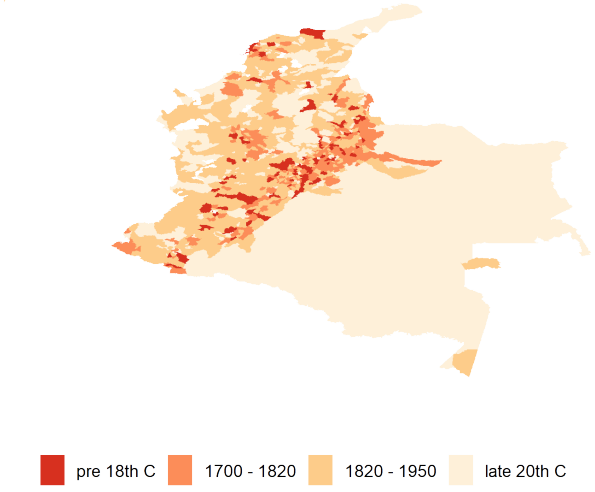
Caribbean macro-region served as the main route to import and export goods and connect the country with the rest of the world. These municipalities then became the main centres of the Colonial activity (Kalmanovitz, 2015).

Figure 2.3: The link between geography and history

(a) Altitude (m above sea level)



(b) Foundation Year



Notes: Panel a shows the average altitude in meters above sea level at the municipal level, panel b shows the foundation year of the municipalities. Sources: Bernard and Zambrano (1993), IGAC.

After 1700, the land frontier expanded from the highlands to lower lands along the Andes, and from the Caribbean to the interior lands of the Magdalena

Valley and the eastern plains. The expansion of the frontier in Colombia is still an ongoing process as the frontier moves to the Amazons and the eastern part of the country to expand mining exploitation and coca cultivation.

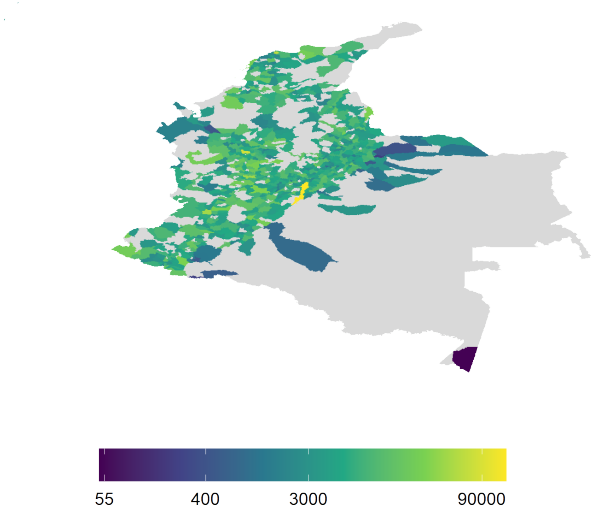
The wide extension of Colombia and the three mountain ranges that cross the country from North to South have presented considerable barriers to the cultural and commercial integration of the regions of the country and between the country and the outside world (Palacios & Safford, 2012; Bushnell, 1996). In consequence, up until the mid-20$^{\text{th}}$ century, the population had been relatively sparse and scattered into small communities. As Fig. 2.4a shows, by 1918 Colombian population was dispersed and divided into small towns, with Bogotá being the only city with almost 150,000 inhabitants, followed by Medellín with almost 80,000 inhabitants. By 1973, the country had experienced a demographic explosion, and most of the population expansion was concentrated in urban centres. Bogotá had almost 2,400,000 inhabitants, Medellín had a population of more than 1,000,000 and Cali and Barranquilla had more than 600,000 people each (Fig. 2.4b). In terms of macroregions, urbanisation was concentrated in the Andean region with some clusters of urbanisation in the Caribbean and in the Pacific.

In terms of economic activity, the industry has been mainly concentrated in the Andean region and industrial production was –and still is- generated by a few sectors (food, beverage and textiles). Also the production of coffee, which was strongly tied to the industrialisation process, has been historically concentrated in the Andean slopes from Santander to Nariño. Other agricultural products such as sugar cane grow in the hinterlands of the Pacific region while most of the grasslands and pasture are found in the Caribbean plains. Gold mining was established in the northern part of the Pacific region and oil exploitation occurs in the Orinoquia region, close to Venezuela. Although regional specialisation was high across the country, this

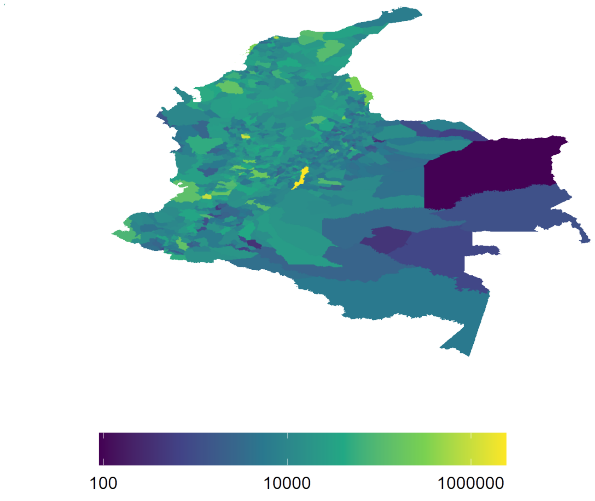
was only for the international markets. The internal market was weak and almost non-existing as most regions produced enough for auto-consumption creating an autarky structure.

Figure 2.4: Population distribution in the 20th century

(a) Population in 1918



(b) Population in 1973

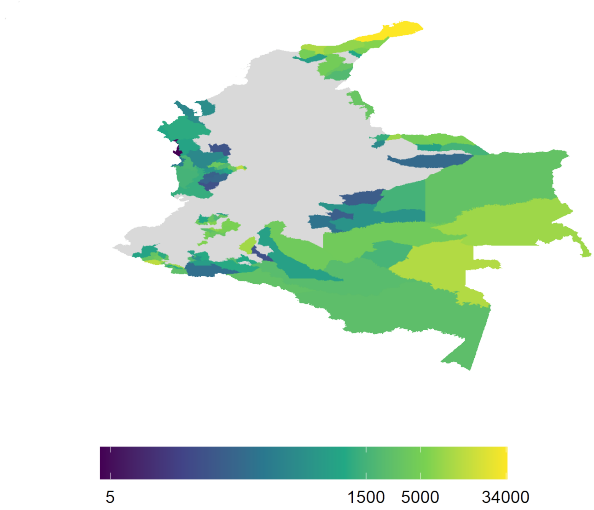


Notes: panel a shows the distribution of the population in 1918 and the scale is in log. The grey area corresponds to territory that had not been occupied or that belong to indigenous communities, as in the case of the Amazonian region. Panel b shows the distribution of the population in 1973 and displays the significant expansion of the frontier after the 1920s. The population in the maps do not include the indigenous population. Sources: National census 1918 and 1973.

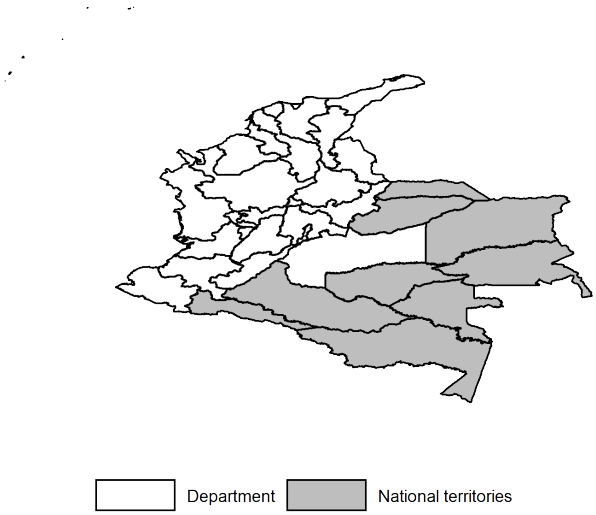
In 1973 the country was divided administratively into 22 departments subdivided into 945 smaller units called municipalities. Additionally, it included 8 national territories which were the more peripheral areas of the country with low levels of population, but that account for almost 50% of the national territory. The division also coincided closely with the regions where indigenous communities have been concentrated since the beginning of the 20^{th} century (Fig. 2.5a).

Figure 2.5: Indigenous population and Administrative division in 1973

(a) Indigenous population in 1973



(b) Administrative division in 1973



Notes: panel a shows the distribution of the indigenous population by 1973 and the scale is in log. The grey area corresponds to municipalities without indigenous communities by 1973. Panel b shows the administrative division of the country during the same year. Several national territories also coincide with indigenous communities. Sources: National census 1973.

2.3 Colombia in the 20th century

As the frontier expanded from the highlands to the lowlands in the 20th century, the country underwent important socioeconomic transformations that affected fertility

patterns across the country. In this section, I discuss some of these transformations, placing special attention on the structural change of the economy (in particular to the rise of industrialisation and urbanisation), the emergence of mass education and modern contraceptive methods, and the potential role of the Government and the Church in the fertility decline.

The expansion of the frontier was translated into an increase in population from 1918 to 1973 which was also accompanied by a rapid urbanisation process. The country went from being mostly rural in 1938 (32% of the population living in urban areas) to mostly urban by 1973 (the urban population was 63%). This agglomeration in the cities significantly differed from the European case during the Industrial Revolution, in which cities were regarded as death traps due to their high levels of mortality caused by overcrowding, lack of sanitation and unhealthy conditions (Komlos, 1998). In Colombia, urban centres enjoyed higher living standards accompanied by higher wages. These characteristics of the industrialised urban centres created a differential gap between rural and urban areas, both in wages and in living standards, that persists today.

Industrialisation took off in 1920, mainly in Antioquia, Valle and Bogotá. This shifted the sectoral composition of the GDP and the employment structure. For example, agricultural production as a percentage of the GDP went down from 46% in 1938 to 25% in 1973.[8] Furthermore, the structural transformation from agricultural sector to the industrial sector happened relatively fast, but there was a faster change towards the services sector as well, which implies that industrialisation was not supported by increases in value-added. As a result by 1980, the sector with higher participation in the GDP was services with 48% of the GDP. In terms of GDP per capita, the economy grew significantly during this period as it increased from 1,808

[8]This structural transformation did not happen across the country and most regions continued to be mostly agricultural with industry and services concentrated in a few cities.

Geary-Khamis international dollars in 1938 to 3,546 in 1973 and 5,938 in 2015. These economic transformations were accompanied by large gains in human development, most importantly in life expectancy (Jaramillo-Echeverri et al., 2019).

Linked to these economic changes, in the 1960s Colombia experienced one of the fastest fertility transitions in the world, in only 25 years the number of children per woman fell from 6.8 to 3.[9] The evolution of this decline has been widely documented in the literature as several surveys were carried out during the second half of the 1960s. A wide range of analyses of these data showed that the fertility decline started around 1964 and by 1985 there was a deceleration of the decline, especially for urban women. Both rural and urban women experienced a decline and the fertility differential between rural and urban narrowed from 4 live births in 1969 to 2 live births in 1986 (Prada-Salas, 1996).

Overall all age groups participated in the decline but it was greater in older women and was possible by having longer birth intervals between births (Batyra, 2016; Parrado, 2000), pointing to the importance of deliberate fertility control. The decline in fertility was tied to the increase in the knowledge, availability and use of contraceptive methods after the mid-1960s. By 1978 90% of women of reproductive age knew of at least one contraceptive method (Prada & Ojeda, 1987); 91% approved of family planning practice in Bogotá in 1974 (Bailey, Measham & Umana, 1976); and by 1969 43% of urban women were practising contraception. Prada and Ojeda (1987) showed that there was an increase in the use of contraceptives, and this increase was more pronounced among rural residents (the proportion went from 15% in 1969 to 53% in 1986). The authors also highlighted that these changes were achieved due to a bigger demand for the pill, female sterilisation and intrauterine devices (IUD). This increase in the demand for more modern contraceptive methods was matched

[9]López Toro (1968) analyses fertility during the intercensal period 1951 - 1964 and shows a high proportion of children between 0 - 5 years old. This evidence suggests that the fertility decline started after 1965, as suggested by Potter, Ordoñez and Measham (1976).

by the increase in supply that arose by the end of 1965 when oral contraceptives became available worldwide and when the Colombian Association for Family Welfare (Profamilia), one of the oldest and largest private family planning organisations in the world, was founded in Colombia.

The effect of family planning provision on the Colombian fertility decline was addressed by Miller (2010). His paper exploits the differences in timing and the geographical patterns of the spread of the family programmes and compares women living in cities that were exposed to the family planning programme at different ages across the country.[10] He finds that family planning explains between 6% and 7% of the fertility decline in these cities between 1964-1993. These results suggest that women living in urban areas were able to achieve their desired number of children using other contraceptive methods such as rhythm or abortion. Abortion was commonly practised in the country during the 1960s, becoming one of the leading causes of hospital admissions for women (Mendoza-Hoyos, 1968). According to Mendoza, more than 60% of abortions happened in women with seven or more births, suggesting that abortion was a common practice to attain (or limit) the desired number of children. Interestingly, induced abortion was mainly present in main cities and the abortion rate per pregnant woman was around 20%, while only 8% of women living in rural areas practised abortion (Requena, 1968).

In the political sphere, the first Colombian plebiscite was held in 1957 being one of the elections with the highest turnout in the history of the country and the first time that women voted. The plebiscite was voted as a result of the peace agreement after a decade long conflict between the two traditional parties: liberals and conservatives. The peace agreement resulted in a political agreement between the two parties to share the power alternating the presidency every four years for four

[10]The spread of Profamilia was fast and cover almost all regions of the country. By 1970 Profamilia had 17 clinics in the Andean, Pacific and Caribbean regions although it did not have any clinics in the regions of Orinoquia or Amazonas.

presidential terms. But it also included the compromise of the government to spend no less than 10% of its budget on public education, following the liberals' ideals of the 1930s. This increase in public expenditure triggered a sharp increase in enrolments in primary, secondary and tertiary education and the increase in enrolment rates was particularly significant for women. By 1970 and for the first time in history, the proportion of women in education was almost half of the total (Ramírez-Giraldo & Téllez-Corredor, 2006).[11]

Several studies looked into the relationship between fertility and education given the remarkable expansion of schooling in the country at the beginning of the 1950s, but they have done so at the aggregated level (Batyra, 2016; Potter & Ordoñez, 1976; Heaton & Forste, 1998). Chapter 4 studies this relationship in depth.

Perhaps a less studied factor in the fertility decline was the role of the Government in setting an agenda to reduce population growth. This strategy started during the Presidency of Alberto Lleras Camargo between 1958-1962, as the Government began to worry about the potential negative consequences of rapid population growth on the development of the country. A more consolidated effort toward a national family planning system came from the Colombian association of medical schools (ASCOFAME) and its Division of Population Studies founded in 1964, by the Minister of Health and Profamilia. By the end of 1969, the Government adopted an official population policy as part of its national development planning (Ott, 1977). The main objective of this population policy was to reduce the rate of population growth by lowering fertility. The strategies to achieve this were divided into: societal and familiar. The first dealt with the need for creating new social roles by increasing the educational level of the population and increasing the participation of women in the labour force. The second acknowledged the fundamental right of the family to choose the number of children they wanted to have. Moreover, the Government

[11]See Chapter 4 for a brief history of education in Colombia. See also Fuentes-Vásquez (2019).

recognised that to guarantee this right, access to family planning information as well as medical services were essential (CONPES, 820, p.36). This population policy was also supported by the incoming Government of Misael Pastrana, who later created the National Population Council in 1970.

These policies were accompanied by a gradual change in the Church's attitude towards family planning. By 1969, a study showed that almost 60% of the priests interviewed considered population growth a serious problem in Colombia and 80% considered that family limitation was necessary (Shea et al., 1971). However, almost 70% thought that contraceptive methods such as the pill or the IDU should never be permitted. The introduction of modern contraceptive methods and its widespread across the country was particularly successful because the responsibility was taken in the first place by a private organisation like Profamilia. Although the pill was available in public hospitals, most of its supply was made by Profamilia.

Despite striking regional differences in Colombia most of the analyses in the literature use aggregated data neglecting the regional character of the decline. As discussed in the previous section, geography in Colombia has, since the pre-colonial period, defined the zones in which the population has settled. This discussion goes beyond mere geographical determinism but acknowledges how geographical barriers led to different colonisation processes and prevented the economic integration of the regions before the late 19^{th} century when investments in infrastructure started to increase. Two main exceptions are the pioneering work of Gutiérrez de Pineda in 1968 and the work of Wills Franco (1976). The ethnographic work of Gutiérrez de Pineda (1968) argues that differences in cultural and historical legacies are reflected in the differences in the process of family formation. The work of Wills Franco (1976) concludes that fertility choices depend on the economic opportunities and on the social norms of each region which affect the net price of the children (Wills Franco, 1976).

However, these works provide limited quantitative empirical evidence. Additionally, the analysis was made at the macro-regional level, while in this thesis I will present empirical evidence at a finer level estimating fertility rates at the municipal and departmental levels.

2.4 Data

My empirical analysis is based on individual-level data from three different censuses. I use the complete census of 1973, which includes detail information at the individual level for all people in the country. Additionally, I use individual-level data from a 10% sample of the census of 1985 from the Integrated Public Use Microdata Series – IPUMS International (Minnesota Population Center, 2019), which is a representative sample at the national and sub-national level. Finally, I use individual-level data from the complete census of 1993. I combine the information on the demographic and socioeconomic characteristics of the individuals with their household information to obtain a detailed dataset. The three censuses report information for all municipalities and departments. With the individual-level information I also create aggregated measures at the municipal level such as urbanisation rates and schooling rates.[12]

Most of the Colombian literature has provided fertility measures with surveys that are representative only at the national level or have not considered a long-run perspective as I do.[13] Furthermore, the vital registration data in Colombia is weak with high levels of under-registration. On the contrary, Colombian censuses are of good quality, and in terms of census coverage, their record is superior to that of most Latin American countries (Vejarano & McCaa, 2002). Colombia has a longstanding

[12]Table 2.6 gives detailed information on the type of census, variables used for individuals and households, the geographical coverage and the source.

[13]One the exceptions is Miller (2010) who uses the census of 1973 and 1993. Similarly, Potter et al. (1976) use samples of the 1973 census to study the fertility decline at the national level but focuses on the 1973 period and does not provide retrospective fertility estimations as I do.

tradition of implementing censuses, and since its Independence in 1820, 18 censuses have taken place in the country, most of them with an interval of 10 years. Therefore, these censuses provide the most complete source of the Colombian population and allow me to explore fertility using detailed socioeconomic variables, such as enrolment in schooling, fertility (measured as children ever born, and children surviving) and housing construction materials.

2.4.1 Census of 1973

The 14^{th} Colombian population census was made on October, 24^{th} 1973, and showed a total population of 21,500,000 people, comparable to that in the United Kingdom in 1851 (around 23 million people). The urban population was held at home during the whole day of the census, while the rural population was enumerated days before (Potter & Ordoñez, 1976). According to the National Agency of Statistics (DANE), the national coverage of the census was 92.8%. A review of the completeness of enumeration of the census Potter and Ordoñez (1976) show that the coverage of the enumeration for women was 94.1% and for men 91.8%.

In this chapter, I provide a detailed analysis of fertility for those living in private households and for the indigenous population as the census used different enumeration forms for people living in private dwellings and for people living in indigenous communities. I treated them separately, as I want to provide a detailed view of the fertility patterns of indigenous populations as these populations have been partially ignored in the demographic literature and they are likely to follow different demographic patterns from the rest of the country. Furthermore, the information at the household and individual levels differ slightly. For example, there is no distinction between rural and urban indigenous groups and the form for indigenous groups asked if they had a radio in the house, but this question is absent from the private household form. Nonetheless, it is not possible with this census to identify ethnicity or race as

this question was not asked in the enumeration forms.[14] Table 2.2 presents the main summary statistics of the census.

Table 2.2: Descriptive statistics: 1973 census

	Private households	Indigenous
Mean age (sd)	21.9 (18.2)	21.4 (17.8)
Mean family size (sd)	7.32 (3.24)	-
Mean numb. children ever born (sd)	3.8 (3.8)	4.1 (3.6)
Urban (%)	62	-
Share with electricity (%)	61	4
Share with sewage (%)	49	4
Share with aqueduct (%)	66	9
Share with primary schooling (%)	58	33
Share with secondary schooling (%)	13	1
Total population	19,867,719	322,788

I do not include in my analysis data for the military forces or institutions such as hospitals.
Source: National Census of 1973, National Agency of Statistics (DANE).

2.4.2 Census of 1985 and 1993

The 15th Colombian population census was made on October, 15th 1985, and showed a total population of 27,837,932 people. It was the first time that enumeration was made *de facto* counting all people living in the house during the time of the census. The urban population was not "immobilised" and the enumeration continued over a period of two weeks starting October 15. In rural areas enumeration continued for one month. According to the National Agency of Statistics (DANE), the national coverage of the census was 92.2%. I use a 10% sample from the Integrated Public Use Microdata Series – IPUMS International and I use the information of all the individuals to calculate fertility rates at the national level given that this sample is representative.

Additionally, I use the information of all women registered in the full census of 1993. In particular I obtain information on the number of children ever born, their

[14]This also means that it is not possible to identify indigenous people living outside the indigenous territories.

age, school level and the municipality of residency. The 16th Colombian population census was made on October, 24th 1993, and showed a total population of 33,109,840 people. The urban population, including settlements with 300 dwellings or more, was "immobilised" and counted on census day. In rural areas, fieldwork continued for a period of six months. According to the National Agency of Statistics (DANE), the national coverage of the census was 88.5%.

Table 2.3: Descriptive statistics: 1985 and 1993 Census

	Census of 1985	Census of 1993
Mean age (sd)	24.6 (18.7)	25.9 (19)
Mean family size (sd)	6.02 (3.18)	5.35 (2.62)
Mean numb. children ever born (sd)	3.06 (3.04)	2.66 (2.98)
Urban (%)	70	72
Share with electricity (%)	80	86
Share with sewage (%)	70	78
Share with aqueduct (%)	71	80
Share with primary schooling (%)	83	86
Share with secondary schooling (%)	38	47
Total population	27,837,932	33,109,840

Source: The data from 1985 comes from IPUMS and the data for 1993 from National Agency of Statistics (DANE).

Regarding the limitations of these sources, censuses can be affected by under enumeration, misreporting of ages, and important differences in group characteristics. For example, age heaping seems to be particularly persistent in people with low levels of education, and under registration is likely to affect children between 0 and 1 year old.[15] This seems to be particularly problematic for the 1985 census but the age heaping bias seems to be stronger in the 1973 census. Additionally, the census does not report the mother's age at birth, which is a key variable to estimate age-specific fertility rates.[16] Section 2.6 explains in detail how I deal with this.

[15]See A'Hearn, Baten and Crayen (2009) for age-heaping and its relationship with literacy and numeracy.

[16]See Reid, Jaadla, Garrett and Schürer (2020) and Timæus (2021) for further limitations of calculating age-specific fertility rates with census data.

2.5 Geographical variation of fertility decline

One of the main advantages of using census data to study fertility decline is that this type of data allows us to measure fertility at different levels of aggregation. In this Chapter, I observe how fertility patterns declined at the municipal and departmental levels in Colombia.[17]

By looking at the number of children ever born to women aged 53 to 63 it is possible to observe fertility patterns before the fertility transition and to uncover the "social forerunners" of this demographic process. Similar to what Livi-Bacci (2017) called "social forerunners" for the European Fertility Project, this group of women are the forerunners of the fertility decline in Colombia as they completed their fertility life before innovative liberal reforms such as mass education in the 1950s, the introduction of modern contraceptive methods or mass media in the mid-1960s. Although these estimations do not include women living in indigenous communities, in section 2.6, I present new estimations of fertility in indigenous populations in the 1950s.[18] Similarly, I can study fertility during the fertility transition to observe how these spots of high and low fertility change over time in the territory. I observe this at the municipal-level, so I cannot uncover class effects, as socioeconomic status varied widely across municipalities. Livi-Bacci (2017) focused mainly on class rather than geography.

To identify these spatial patterns, I calculate the average number of children ever born at the municipal level, focusing on the municipality of residence of women that reported having at least one child and are aged 53 to 63 in the 1973 and 1993 censuses. I focus on these women as this group allows me to observe women that are unlikely to have another birth. These women allow me to observe fertility patterns

[17] In Chapter 2 I also focus on the municipal level and Chapter 3 exploits the individual-level characteristic of the censuses to observe fertility changes at a finer level.

[18] Most of the fertility analysis of the country do not include information of the indigenous population but the 1973 and the 1993 censuses allow for a representative sample to observe indigenous women. In this chapter I focus on the information from the 1973 census.

before the fertility transition. Women aged 53 to 63 years old in the 1973 census, were born between 1910-1920, a period when the country was changing slowly, while women aged 53 to 63 years old in the 1993 census, were born between 1930-1940, a decade in which the country saw numerous reforms that aimed to promote gender equality. I include only mothers because this allows me to study how the number of children a woman had (also known as the extensive margin) changed before and during the fertility transition. In Chapter 4 I provide a more detailed analysis of the intensive margin.[19]

The main analysis is based on the place of residency of the woman rather than the place of birth because the administrative division of the country changed considerably from 1910 to 1973. Additionally, there are several places in which the proportion of non-migrant women from this cohort is low and municipalities like Arauquita or Calamar reported zero non-migrant women in the 1973 census. Although 58% of the women born between 1910 and 1920 were born in a different municipality, most of these women had been living in that municipality for more than 36 years.[20] Additionally, 67% of the women that migrated moved to a municipality within the same department. This implies that most women moved before completing their fertility life and they moved inside a similar geographical context. The results using only non-migrant women are reported in Appendix Fig. 2.16.[21]

[19]Although the contribution of childlessness is an important part of declining fertility, the spatial patterns of fertility do not change substantially if one focuses on all women. The average number of children ever born will be lower if all women are included in the sample, but the geographical patterns are overall similar, as shown in Fig. 2.15.

[20]This information comes from the census question "How long have you been living in this municipality?".

[21]Looking at migrant assimilation and how fertility patterns changed for migrant women in the 1920s is beyond the scope of this thesis, but it is definitely a good research avenue to pursue in the future. See Daudin et al. (2019) for the French case.

2.5.1 Local Indicators of Spatial Association – LISA

To determine if the geographical patterns have a significant positive or negative association with each other I measure the Spatial autocorrelation among municipalities. Spatial autocorrelation is a statistic that measures the degree of dependency among observations in geographical space and provides a measure of the correlation between a variable in different locations in space. This statistic can result in three different structures: positive, negative and zero autocorrelation. Positive autocorrelation implies that similar places are sharing similar values of the variable of interest, in other words, observations with high (or low) values of the variable of interest tend to be clustered in space. Negative autocorrelation, on the contrary, implies that places tend to be surrounded by neighbours having dissimilar values, a checker-board, for example, with white regions presenting positive values and black negative ones. The third outcome is zero autocorrelation and indicates that the location of data points is not related in the space i.e. the data are randomly distributed in the space.

To detect significant spatial autocorrelation in the data, I use the Local Indicators of Spatial Association – LISA developed by Anselin (1995) to identify those locations or sets of contiguous locations that are significantly positively correlated and those that can be characterised as outliers (negative correlation). The LISA test is based on the null hypothesis that the variable of interest, in this case number of children born, is randomly distributed among the entities in the study area. When the P-value is statistically significant, then the null hypothesis can be rejected, and positive clusters or negative associations would be established. Places that had a significant spatial autocorrelation are defined as high-high clusters (in red), which are places where the variable of interest takes high values and is surrounded by places with high values; low-low clusters (in blue) which include places with low values with neighbours with low values, high-low clusters (in light red) which are places with high

values of the variable surrounded by places with low values and low-high clusters (in light blue) which are places with low values of the variable, surrounded by places with high values of the variable of interest.

Additionally, one crucial issue in spatial econometrics is how to define closeness in space. Typically, spatial analysis uses the spatial weight matrix that delineates the neighbours for each location by weighting the intensity of the relationship among a pair of spatial units. There are two different approaches to creating neighbour's weight matrices. The Rook matrix is based on contiguity of order 1 or higher and the Queen is based on distance. I use the Rook matrix which is based on contiguous boundaries, instead of establishing a distance that makes sense for the whole country.[22] A key point after defining the spatial weight matrix is to compute weighted averages in which more weight is placed on nearby observations than on distant observations. In my analysis, I choose a 5-order continuity, which means that I consider as a neighbour all the neighbours of neighbours up to 5 degrees, placing more weight on those that are closer in the distance to the municipality.

2.5.2 Fertility before the decline: the 1910-1920 cohort

The 1910-1920 cohort is composed of all women with completed fertility in 1973, which allows me to measure fertility patterns before the introduction of liberal reforms after the 1930s (mass education, modern contraceptive methods and mass media).

According to the 1973 census, the average number of children per woman in the 1910-1920 cohort was around 7. But the data also shows an important geographical variation. While women living in Ospina, Nariño, a small rural town located near the frontier with Ecuador, had on average 3 to 4 children, women in Betulia, Antioquia, also a small town 50 km away from Medellín, had 9 children. As Livi-

[22]Municipalities in Colombia have very different sizes, in particular municipalities in the Amazonia and Orinoquia regions. This makes it difficult to determine neighbourhoods based on distance.

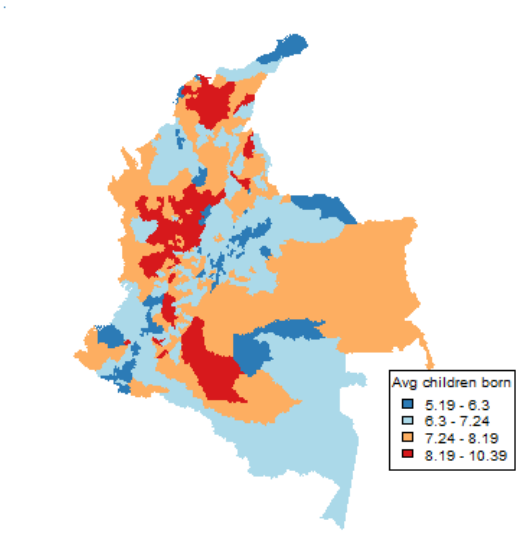
Bacci (2017) argues for the European case, in Colombia there were also some regional pockets of low fertility before the fertility transition.

The geographical distribution of the average number of children per woman is shown in Fig. 2.6a and confirms the existence of different patterns of fertility in the country. Overall the map shows that most of the pre-transition low fertility municipalities are found over the Andean mountain ranges in the Andean region, with pockets of low fertility in the Caribbean and the Amazonian regions. Municipalities in the Pacific, some parts of the Caribbean and the Orinoquia regions appear as high-fertility municipalities. On average, a woman located in a pre-transition low fertility municipality had 1 fewer child than the average woman from the same cohort, while a woman located in a pre-transition high-fertility municipality had 2 children more. The results from the LISA tests reported in Fig. 2.6b uncover significant geographical patterns of pre-transition low and high fertility. The results show that pre-transition low fertility clusters are found in municipalities of Cundinamarca, Boyacá, Santander, Cauca and Nariño, while pre-transition high fertility clusters are found in the departments of Antioquia, Bolívar, Caldas, Caqueta, Magdalena, Chocó, and the northern part of Valle del Cauca.[23]

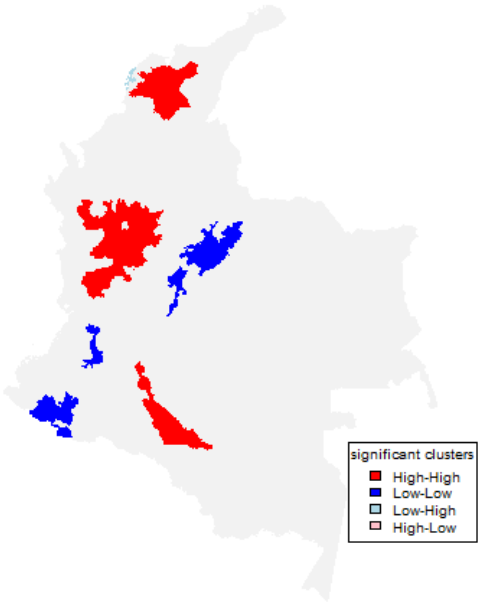
[23]Appendix Table 2.11 and appendix Table 2.8 present the complete list of the municipalities in high and low clusters by the department.

Figure 2.6: Spatial association in completed fertility in 1973

(a) Municipality effect



(b) Significant spatial association



Notes: Fig. 2.6a plots average number of children at the municipal level. The clusters are defined by the effect of the municipality of residence on the number of children ever born to women born between 1910-1920. Fig. 2.6b shows the results of the LISA test. Sources: Authors' calculations based on 1973 Census data.

Pre-transition high fertility clusters share two main historical and geographical characteristics: these municipalities witnessed the expansion of the frontier at some point after the mid-19th century, and they are mostly located in lowlands. As Table 2.4 shows the average municipality in the high clusters has 1,313 m elevation and was founded around 1861. The northern cluster is composed of the municipalities of Bolívar and Magdalena which belong to the Caribbean macro-region which has similar cultural characteristics and ethnic compositions. In this macro-region, we also observe Cartagena de Indias as an outlier, which implies that the city has low fertility in comparison to its neighbours. It is noticeable that Cartagena de Indias, which was the main port during the colonial times appears as the outlier while Barranquilla, one of the 4 most prominent urban centres and the main port in 1973, exhibits higher fertility. Once again, this result corroborates the weak link between socioeconomic conditions and fertility and suggests that other factors, such as cultural and historical legacies should be taken into account to fully understand the fertility transition.

Another high-fertility region is formed by the municipalities of Antioquia, Caldas, Chocó and the Northern part of Valle de Cauca. Most of the municipalities of this cluster are located in the lowlands of the Andean macro-region although it also includes some municipalities of the Pacific macro-region. Only 4 of these municipalities were founded before 1600 while most of them were founded after the 1860s, as a result of an expansion of the frontier from Antioquia to the Northern part of Valle del Cauca and Chocó during the late 18th and 19th centuries. The departments of Caldas, Risaralda and Quindio were founded and populated by people from Antioquia, as well as some towns in Valle del Cauca (Parson, 1950).

The last high-fertility region appears in the Amazonian macro-region in the department of Caqueta which is also a territory that experienced an expansion of the frontier during the 1940s.

Pre-transition low fertility clusters also share similar historical and geographical characteristics. Most of the municipalities in this cluster are found along the Andean mountain range (on average these municipalities are located over 2,500 m above sea level) and all of them are close to important colonial centres and urban areas in the 20th century. On average, municipalities that appear as low-fertility clusters were founded in 1771 and 81% of their population was living in the urban centre of the municipality in 1973. Nonetheless, several municipalities, in Boyacá, Nariño and Cundinamarca have an urbanisation rate of around 15%, which makes the relationship between urbanisation and low fertility less clear.

The first low-fertility cluster is formed by municipalities in the departments of Cundinamarca, Boyacá and Santander. The cluster is located in the *Altiplano Cundiboyacense* which corresponds to the ancient territory of the Muiscas, the indigenous culture that formed one of the best-organised confederations of tribes on the South American continent before the arrival of the Spanish. The cluster includes several important colonial centres such as Bogotá, Tunja and Villa de Leyva that were founded around 1540 and where women born between 1910-1920 had on average 4 children. The cluster also includes several small towns with low urbanisation rates but with high access to public services given its closeness to the capital.

The next low-fertility area was developed around two main cities founded in the 1530s by Sebastian de Belalcazar: Cali, the capital city of the department Valle del Cauca and Popayan, the capital city of Cauca. The cluster also includes small towns such as Puerto Tejada and Santander de Quilichao where most of the population were agricultural labourers by 1973. In the Southern part of Colombia, close to Ecuador, the municipalities in the department of Nariño from another cluster. Once again, the cluster is formed around an important colonial centre, Pasto, also founded by Sebastian de Belalcazar in 1539. Surrounding Pasto there are several

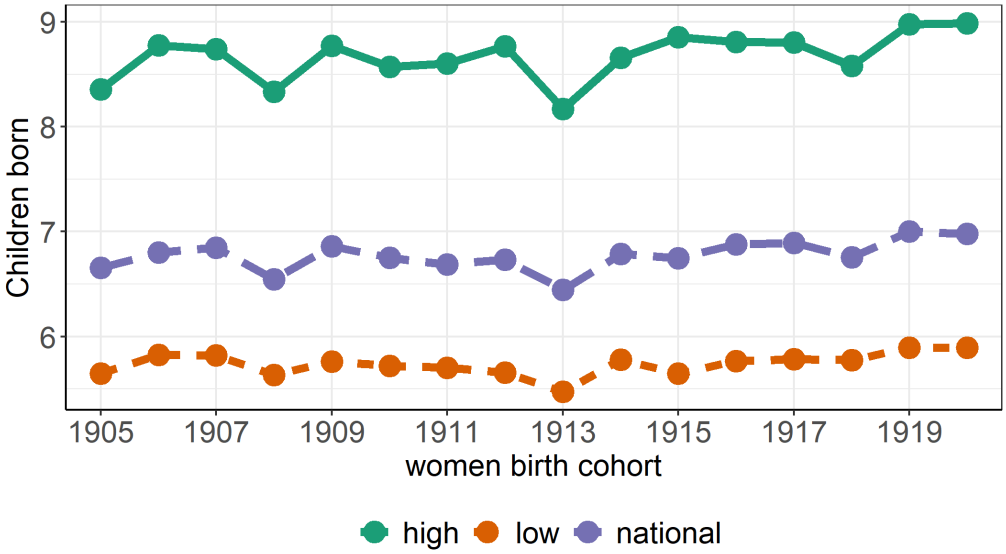
small towns where traditionally the main economic activity is agriculture and the most important crops, even today, are potatoes, cocoa, corn, and wheat.

Generally speaking, the geographical patterns uncovered by the cluster analysis reflect differences in historical legacies. This new empirical evidence suggests that the relationship between industrialisation and fertility patterns during the fertility transition in the Colombian was not strong.

The differences between the clusters in fertility patterns before the onset of the fertility transition are shown in Fig. 2.7 and confirm that there were persistent differences in completed fertility for women born between 1905 and 1920 across the clusters.[24] On average, women in low-fertility regions had almost 2 children fewer in comparison to women in high-fertility areas, and the gap was stable throughout the different cohorts. These results confirm that before the fertility transition the clusters were stable and that low-fertility clusters did not experience earlier fertility decline despite having smaller fertility levels.

[24]Given that we cannot provide estimations of the TFR for years before 1958 Fig. 2.7 displays the number of children ever born for women born between 1905 and 1920 as a way of observing fertility patterns before the fertility transition.

Figure 2.7: Pre-transition completed fertility



Notes: The graph shows the average number of children ever born by cohort from 1905 to 1920. By 1973 these women had completed their fertility life and therefore are a good approximation of the fertility levels in the clusters before the fertility transition. Dips for those reporting ages ending 0 and 5 are likely to be linked to age reporting. Sources: 1973 Census data.

By 1973 the clusters had noticeable differences in their levels of development when measured by urbanisation and access to public services (see Table 2.4). A higher proportion of women in the pre-transition low fertility areas had completed secondary schooling and this difference is partially reflected in the percentage of women that appear registered as paid employees, although in both clusters the female labour force participation was very low.[25] Interestingly, the share of married women and the sex ratio was not very different in the clusters.

[25]When looking at the type of employment of women older than 15 years old, 72.4% do not report any occupation in low fertility clusters while 89.1% do not report occupation in high fertility clusters.

Table 2.4: Summary statistics of the clusters, 1973

	Low clusters	**High clusters**
Family size	7	8
Children born (sd)	3.2 (3.3)	4 (4.4)
Urban (%)	81	43
Share with electricity (%)	80	47
Share with aqueduct (%)	79	64
Women with secondary schooling (%)	19	8
Women payed employees (%)	15	6
Women married (%)	45	50
Sex-ratio (men/women) age 15 to 50	0.85	0.95
Altitude (m)	2,492	1,131
Year of foundation	1771	1861
Total women older than 15 years old	**1,558,072**	**597,860**

Notes: urbanisation and access to public services are based on the complete population of the municipalities. Women enrolled in secondary education and in the labour market are based on women older than 15 years old. Source: Authors' calculations based on 1973 Census data.

2.5.3 Fertility during the decline: the 1930-1940 cohort

It has been shown that geographical differences in fertility levels, before the fertility transition, were strongly linked to differences in historical legacies. To observe how much spatial patterns changed during the fertility transition I focus on women born between 1930 and 1940 that appear in the 1993 census. These women were born after the liberalisation of education for women and the elimination of the marriage bar in the early 1930s. Additionally, these women were exposed to modern contraceptive methods such as the pill and started their fertility life during the onset of the fertility transition. Following the same approach as before, I estimate the average number of children ever born at the municipal level and I use the LISA test to find significant spatial auto-correlation in the data.

The average number of children ever born to this cohort of women was around 5, confirming that fertility declined with respect to the previous cohort, but

again the data reveals considerable municipal variation although the spatial patterns seem to have changed. Women living in the municipality of Roberto Payan in Nariño, a small town 250 km away from Pasto and close to the Pacific Ocean, had around 9 children per woman, while women living in Riosucio Caldas, a town founded in 1819 during the expansion of the frontier from Antioquia to the Southern part of the country, had fewer than 3 children. The geographical distribution of the number of children ever born is shown in Fig. 2.8a. Unlike the results for the 1910-1920 cohort in Fig. 2.6a, high-fertility municipalities are found all across the country, even in high-altitude regions over the mountain Andean range in the department of Antioquia, as well as in some parts of the coastal regions and the Amazonian and Orinoquia macroregions. For their part, low-fertility municipalities are found along the Pacific region, except for the municipality Alto Baudo in Chocó, over the Andean mountain range in the departments of Risaralda and Quindio and some municipalities in Nariño. When Fig. 2.6a and Fig. 2.8a are compared, the clearest patterns are that several low-fertility municipalities appear as such in both cohorts, mainly over the Andean mountain range from Nariño to Bogotá, while high-fertility areas expanded towards neighbouring municipalities, especially in Caqueta and towards Magdalena.

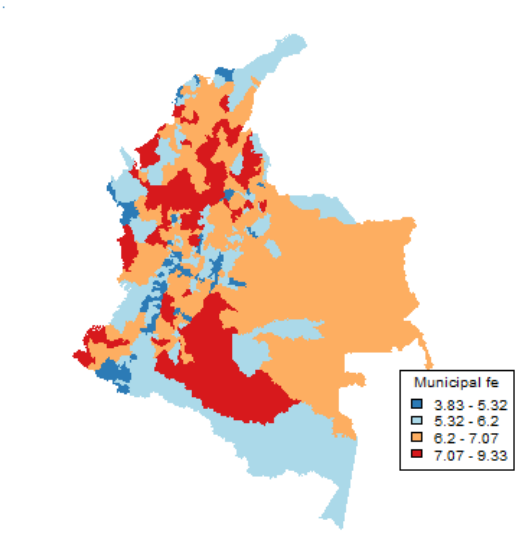
The results from the LISA test reported in Fig. 2.8b confirm that the patterns differed in comparison to the clusters found before the fertility transition. High fertility clusters shifted their location and expanded to new areas. The cluster in the northern part of Colombia moved from the coastal part of the departments of Bolívar and Magdalena towards the interior of the country moving into the department of Cesar. The cluster that was formed by municipalities of Antioquia, Caldas, Chocó and the Northern part of Valle de Cauca, moved away from the Pacific Coast into the North East of Antioquia, creating an almost unique region with the cluster made by the municipalities in Cesar and Magdalena. It is noticeable that almost all municipalities in the department of Chocó, one of the most underdeveloped areas

of the country by 1993, show low levels of fertility and that the municipalities of Bahia Solano, Bojaya and Jurado limit with the department of Antioquia, appear as outliers, that is, municipalities with low fertility surrounded by municipalities with high fertility. Lastly, the cluster of Caqueta, which previously was formed by the municipalities of Puerto Rico and Cartagena del Chaira, expanded, covering almost all the municipalities of the department, except Florencia, the capital city.

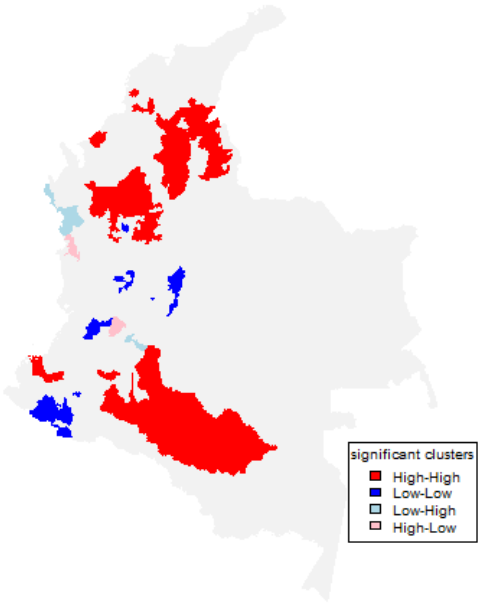
Low-fertility clusters also changed, except for the cluster formed by the municipalities in Nariño, Cundinamarca and the city of Cali, which suggest that in these places low fertility has persisted across generations. The low-fertility clusters that were previously found in the municipalities of Boyacá and Cauca do not appear to be regions with significantly lower fertility by 1993. New dispersed low-fertility clusters appeared in the departments of Risaralda, Quindio and Antioquia, in particular in the capital cities of these departments and the municipalities that surround them, which were municipalities with relatively high fertility before the fertility transition. In particular, the municipalities of Envigado, Itagui, Sabaneta and La Estrella in the department of Antioquia were part of a cluster of high fertility for the previous cohort, indicating a significant change in the fertility patterns in these regions during the fertility decline.

Figure 2.8: Spatial association in completed fertility in 1993

(a) Municipality effect



(b) Significant spatial association



Notes: Fig. 2.6a plots the average number of children by municipality. The clusters are defined by the effect of the municipality of residence on the number of children ever born to women born between 1930-1940. Fig. 2.6b shows the results of the LISA test. Sources: Authors' calculations based on 1993 Census data.

Table 2.5: Summary statistics of the clusters, 1993

	Low clusters	**High clusters**
Children born (sd)	2.3 (2.6)	3.4 (3.8)
Urban (%)	90	32
Women with secondary schooling (%)	19	4
Sex-ratio (men/women) age 15 to 50	0.87	1.06
Altitude (m)	2,326	982
Year of foundation	1807	1897
Total women older than 15 years old	**3,837,597**	**445,440**

Notes: urbanisation and access to public services are based on the complete population of the municipalities. Women enrolled in secondary education and in the labour market are based on women older than 15 years old. Source: Authors' calculations based on 1993 Census data.

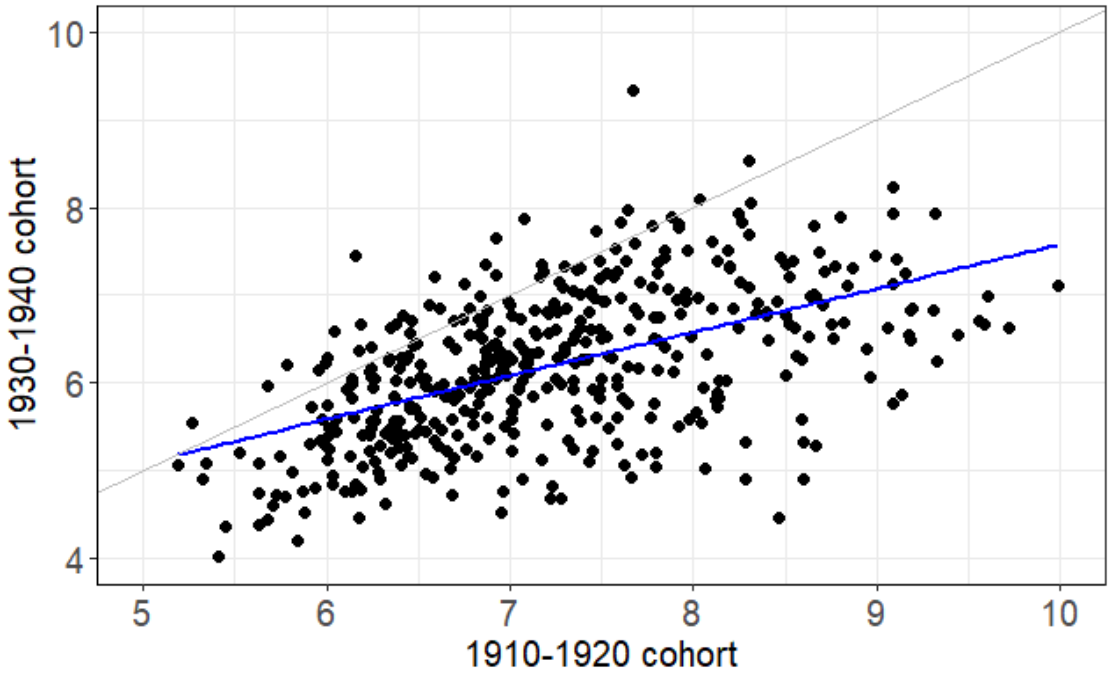
Largely, before and during the fertility transition high fertility clusters were located in municipalities that were founded after the mid-19th century, with an altitude below the 1,100 m above sea level and where most of its population was living in the rural areas, even by 1993, when the national urbanisation rate was 72%. This meant that women had fewer opportunities to enrol in secondary education, despite the laws that permitted them to do so. But in any case, these high-fertility areas also experienced a reduction in the number of children ever born, and by 1993 women living in these high-fertility regions had on average 1 child fewer in comparison to women living in high-fertility areas in 1973.

For their part, low fertility areas were concentrated before and during the fertility decline in municipalities around the 2,400 m of elevation. But new zones of low fertility emerged during the fertility decline in cities such as Pereira and Armenia that were founded after the mid-19th century, indicating that the relationship between fertility and historical legacies started to weaken while the importance of living in a well connected urban centre became more prominent. Finally, it is also worth noticing that the number of women living in low-fertility areas almost doubled from 1973 to 1993; on the contrary, it decreased in the high-fertility areas during the same

period. This implies that low-fertility areas became more populous across the territory pushing the fertility decline as an aggregated effect. Overall, low fertility areas in 1993 seem to be related more to urbanisation and the concentration of population rather than the spread of low fertility into new places. By 1993 most women were living in places where low fertility was the most accepted social norm, and at the same time, places with higher fertility also saw a reduction in the number of children ever born.

At the municipal level, fertility declined across the country. Fig. 2.9 plots the association between the average number of children ever born in the 1910-1920 cohort and in the 1930-1940 cohort at the municipal level. The estimated association between the two cohorts is 0.82, which implies a strong association between both cohorts. The coefficient of 0.58 suggests that there is a positive relationship between the number of children in 1910-1920 and the average number of children in 1930-1940. This means that municipalities with higher fertility before the fertility transition (as measured by the number of children in 1910-1920) tend to have higher fertility 20 years later (as measured by the average number of children in 1930-1940). Similarly, municipalities with lower fertility before the fertility transition tend to have lower fertility 20 years later. Overall, this suggests that there is some level of persistence in fertility levels across municipalities. However, the coefficient also reflects the generalised decline of fertility in the country. In almost all municipalities (those under the grey line), fertility was lower in the later cohort.

Figure 2.9: Association between cohorts



Notes: The graph shows the correlation between the average number of children ever born in the 1910-1920 cohort and in the 1930-1940 cohort. Each point represents a municipality. The blue line is a linear fitted line and the predicted equation is $y = 3.6 + 0.58x$ with an R-squared of 0.82. The grey line corresponds to the line of perfect association. Sources: 1973 and 1993 Census data.

2.6 Fertility Rates 1958-1990

The previous section uncovers geographical patterns in fertility decisions before and during the fertility transition at the municipal level. But how fast was the decline across the country? Which region or departments preceded the fertility decline and which lagged? To answer these questions, I estimate the Total Fertility Rate at different levels of aggregation from 1958 to 1990 using the Own Child Method (OCM).

2.6.1 Own Child Method

Given the poor quality of vital registration for this period in Colombia, I construct Age-Specific Fertility Rates (ASFR) and Total Fertility Rates (TFR) at the depart-

mental level from the complete 1973 census using the Own Child Method (OCM).[26] The OCM was developed in the 1960s to calculate fertility rates when birth registration data is incomplete or unavailable, or when the mother's age at birth of the child is not registered, as is the case in the census (Grabill & Cho, 1965; Cho, Retherford & Choe, 1986). This method has been widely used to estimate fertility in historical censuses and for developing countries. For example, Reid et al. (2020) estimate fertility rates for England and Wales in 1911; Dubuc (2009) estimate fertility by ethnic and religious groups in the UK; Avery, St. Clair, Levin and Hill (2013) present a cross-county comparison, and Dribe and Scalone (2014) use this method for Sweden between 1880-1970.

The method is based on two key elements of the census: the recorded age of people in the household, and, when available, the relationship of each member of the household to the household head. One advantage of the 1973 census over other data sets is its detailed information about the relationship to the household head, which allowed me to match children with their biological mothers, and to a rich set of socioeconomic variables for both the household and the individual.[27] I pair a potential child with a potential mother following different rules depending on the relationship of both to the head of the household, their age and the fertility reported by the potential mother. All persons in the sample are eligible to receive a mother's link if they have a specified relation to the head. Additionally, all women over age 15 are eligible to be a mother if they do not explicitly report having no children. As the census reports the number of children born and surviving for each woman, I use these variables as a cap to the number of children I assign to each woman.

The set of rules to link a potential mother with a potential child are defined

[26]For the national and cluster-level estimations of the TFR I also use the 1985 census and the 1993 census.

[27]The 1973 census provides 10 different categories for relation to the head of household: head, spouse, child, parent, grandchild, another relative, domestic employee, lodger, other non-relative and unknown.

following Sobek and Kennedy (2009). The Integrated Public Use Microdata Series - IPUMS used these rules for their samples (including the 1973 sample of Colombia) to create links between partners, and parents and children. The set of rules is explained in Appendix Table 2.7.

After obtaining the links between mother and children, I calculate the mother's age at birth as the difference between her age and her child's age. Following Reid et al. (2020), I calculate single years ASFR to compute the TFR for 1958 to 1990, as shown in Eq. (2.1) and Eq. (2.2). These rates are calculated using children between 0 to 14 years old, and mothers between 15 and 78 years old in the census year. The selected age of children avoids the common issue of not matching children that are not living with their parents.[28]

$$ASFR_a, y = \frac{\text{Adjusted number of children born to women age } a \text{ in year } y}{\text{Adjusted number of women age } a \text{ in year } y} \tag{2.1}$$

$$TFR_y = \sum_{age=15}^{64} ASFR_{a,y} \tag{2.2}$$

I define the ASFRs of each year as the ratio between all children born in the year y (children age 0 in the census were born between 1972 and 1973 and children age 14 were born between 1958 and 1959) to women aged 15 to 64 at each year y. I calculate the Total Fertility Rate as the sum of the ASFR at each year.

As the census would not report the number of women that have died in previous years, I adjust the number of women at each year by the probability of surviving of the mother using single-year life tables. Similarly, as the census only

[28] Also as discussed below the possibility of orphanhood increases with the age of the children and the mother.

reports the number of children that are alive in the year of the census, I adjust the number of children at each age with the life tables. To control for potential differences in mortality regimes by departments, I use life-tables at the departmental level when I estimate the TFR for municipalities and departments. For national estimations I use the national life-table of the same decade of the census.[29] Given that I am estimating the ASFR based only on those pairs of mother-child that I was able to link, I adjusted the number of children for the proportion of children at each age that I did not link: in other words, I adjust for the proportion of matched children.[30] Recently Timæus (2021) demonstrated that this adjustment could lead to a biased estimation of the fertility patterns overestimating fertility at younger ages and underestimating fertility at older ages. However, Timæus (2021) admits that the bias in estimates of total fertility rate is smaller as the errors for women of different ages more or less cancel each other out. Additionally, he highlights that this concern affects particularly estimations that are done using the data on children aged 15 or more. I provide estimations of Age-Specific and Total Fertility Rates looking at children 14 years old or younger, which minimises this concern.

2.6.2 Colombia Total Fertility Rate: 1958-1990

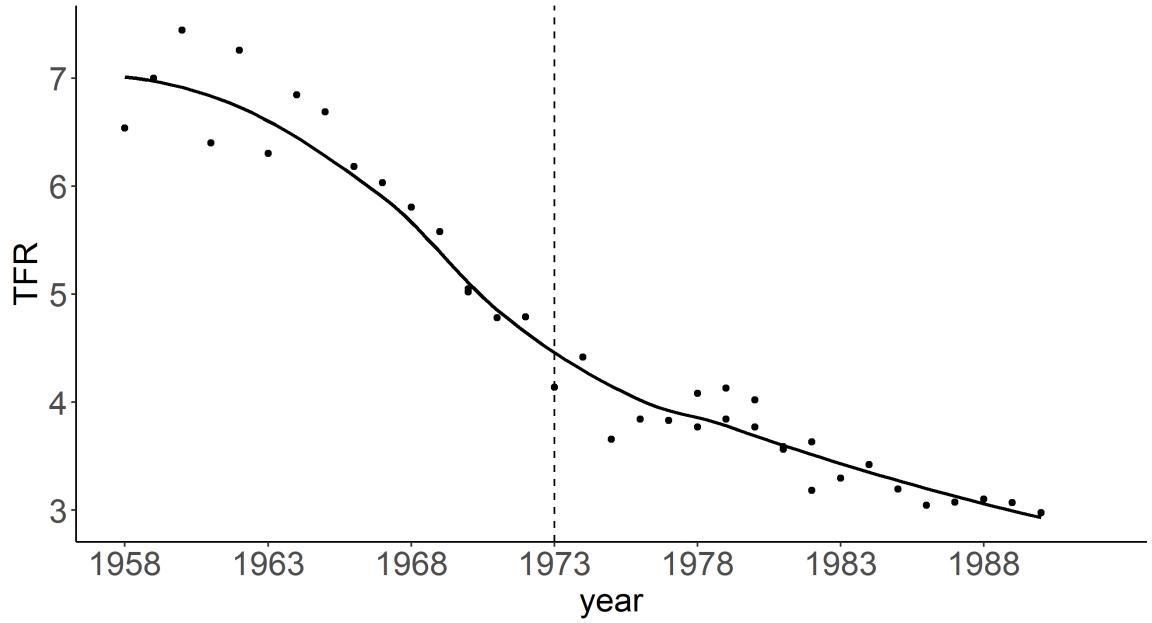
Fig. 2.10 shows the Total Fertility Rate for Colombia for the period 1958 to 1990. The estimations are made using the complete census of 1973 and the IPUMS-international samples of 1985 and 1993. The results show a clear drop from around 7 children per woman to less than 3. These results confirm that fertility declined rapidly around the mid-1960s until the late 1970s and then there is a deceleration of the decline after 1985 (Potter & Ordoñez, 1976; López Toro, 1968; Prada & Ojeda, 1987). However, some

[29]The life-tables come from the National Agency of Statistics (DANE). Most of them were an abridged table and to obtain single-year survivorship probabilities I use the `MORTPAK` package from the United Nations. For details about the package and other forms of expanding an abridged life table see Kostaki and Panousis (2001) and Heligman and Pollard (1980). I use life-tables from 1970, 1980 and 1990.

[30]On average I linked 87% of children age 0 and 81% of children between 5 to 14 years old.

caveats about these results have to be discussed. The fluctuations from 1958 to 1963 can indicate age heaping among older children reported in the census, especially those aged 10 and 12. Additionally, the values for the 2 years previous to the censuses are considerably smaller in comparison to what the overall trend suggests. This could be the result of misreporting of the age of children that are about to turn one or two years old. Another possible explanation is that there are unrecorded births and under-enumeration of zero-years-old (Reid et al., 2020). Therefore, in the following section, I remove these two observations and I focus my analysis on the TFR from 1958 to 1971, where the steeper decline occurred. Also, when estimations come from different censuses, I keep both measures, as in the years 1970, and 1978 to 1982, and the loess curve averages these point estimates given that the estimations do not perfectly overlap because as explained by (Reid et al., 2020), the censuses can suffer from differential under-enumeration by age and census year. The loess smooth curve average both point estimates.

Figure 2.10: TFR 1958-1990, Colombia



Note: Calculations using the OCM. Women between 15 to 78 years old and children between 0 to 14 years old. See Section 2.6 for a description of the methodology. The black dotted line indicates the year of the 1973 census. The black line is the loess smooth curve and the black dots represent the TFR estimation for each year. Sources: Authors' calculations based on 1973 full Census data, and 1985 and 1993 Census samples from IPUMS-International.

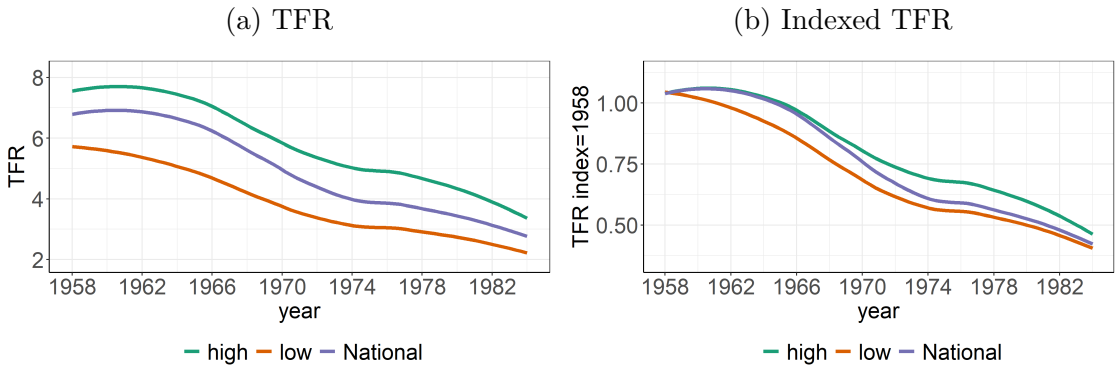
To understand the evolution of fertility in the clusters defined with women born between 1910-1920, and to test if fertility declined rapidly and simultaneously even in places with traditionally different levels of fertility, I compute The Total Fertility Rates from 1958 to 1984 at the cluster level.[31] I use the clusters defined using the 1910-1920 cohort, that is the clusters that show the fertility patterns before the fertility decline. Additionally, I present the TFR indexed to 1958 to observe the speed of the decline.

The TFR in Fig. 2.11a corroborates that by 1958 there were significant differences in fertility levels between the clusters and there was a gap of almost 2 children between the clusters. Additionally, it shows that fertility declined consistently at the

[31]I compute TFR at the municipal level from 1958 to 1984 using women between 15 to 78 years old and children between 0 to 14 years old following the OCM. Given that I do not have life-tables at the municipal level, the mortality adjustments are made at the departmental level. The information comes from the complete census of 1973 and the 10% sample of 1985.

national level, but also in pre-transition high and low fertility clusters. The indexed TFR in Fig. 2.11b confirms that the national drop in the Total Fertility Rate started around 1962 and accelerated in 1964, and this seems to be the case also for the pre-transition high fertility clusters, which follow a very similar pattern as the national until 1970. Fertility in pre-transition low fertility clusters declines at a consistent rate after 1958 decelerating after 1974 when it matches the national trajectory. All in all, during these 25 years the initial gap between high-fertility and low-fertility clusters narrowed to one child by 1983.

Figure 2.11: TFR 1958-1985, geographical clusters



Notes: Fig. 2.11a shows the lowess smooth curve of the Total Fertility Rate. The TFR is calculated using women between 15 to 78 years old and children between 0 to 14 years old following the OCM. Fig. 2.11b shows the lowess smooth curve of the indexed TFR. Sources: Authors' calculations based on 1973 Census data and a 10% sample of the 1985 Census from IPUMS.

These results show that at the beginning of the 1960s, fertility declined fast in places where women traditionally had low fertility as well as in places where women had higher fertility. Although the clusters did not achieve convergence in fertility rates by 1983, the TFR halved in both clusters and at the national level in only 25 years. Overall, this is a remarkable finding as it shows that after 1964 fertility declined at a similar rate in zones with different historical backgrounds suggesting perhaps a process of diffusion of social norms.

2.6.3 Departmental Fertility Rate: 1958-1971

To provide a more granular analysis I calculate the TFR at the departmental level. The results confirm that the decline was fast and simultaneous across the country but with considerable variation in levels both in 1958 and 1971. Fertility levels between departments differed by as much as three children per woman in 1958, but almost all regions show a similar pattern of decline. In 1958, the territory with the highest fertility rate was Putumayo, in the Amazonian macro region, with a TFR of 8.5 while Valle del Cauca, which geographically belongs to both the Andean and Pacific regions, was the department with the lowest fertility rate (5.5). By 1970 the regional differences did not disappear, but fertility fell from 8.5 children to 6.4 in only 12 years in Putumayo and the change was similar in Valle del Cauca, where fertility rates dropped from 5 to 3.3.

All the departments in the Caribbean macro-region exhibited a steady fertility decline after 1960 while most departments in the Pacific macro-region show relatively low fertility in comparison to other macro-regions. The exception is the department of Choco which appears as the department with one of the highest TFR in the country. But as argued before, these departments also underwent a fertility decline after 1965. In the Andean macro-region it is noticeable that the department of Boyacá did not experience a sharp decline in fertility despite starting from a relatively low level. On the contrary, the TFR shows a relative increase in fertility up to 1965, when fertility started decreasing and by 1970 the TFR was similar to that of 1958. The rest of the departments in the Andean region show a steady decline despite differences in initial levels.

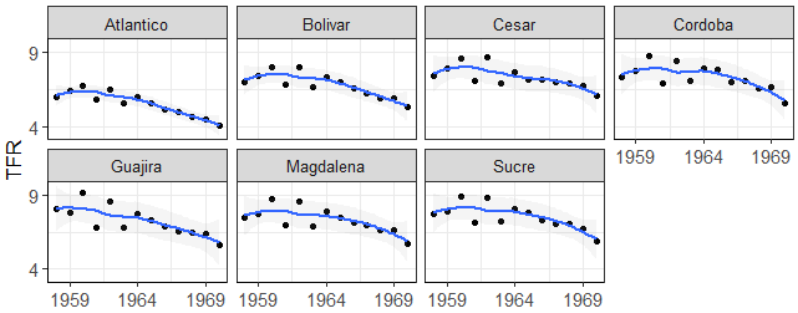
Although by 1958 most of the departments in the Amazonian and Orinoquia macro-region exhibit a high fertility rate, with a TFR close to 9 children per woman, the territory of Vaupes and the islands of San Andres and Providencia had a TFR

lower than 5. The low fertility rate of Vaupes in 1958 might be explained by the scarce population that was living in the area, as Vaupes has been historically the least populated department in the country. However, the result could also suggest under-reporting of 14 years old or early age at leaving home. The differences in fertility levels are also noticeable in 1970. Guainia and Vaupes saw an increase in the TFR although with an irregular pattern. Both territories were scarcely populated until the late 1950s and their fertility patterns from 1958 to 1970 could be explained by the increase in the inhabitants of the area that resulted from migratory movements.

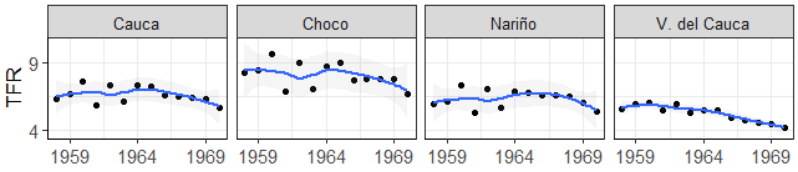
Similar to the national TFR, the estimations are noisier for some departments and territories in the years 1960 to 1963 which could reveal age-heaping (and numeracy) problems or age misstatement among children, in particular in departments that are located in peripheral areas such as Chocó, Córdoba, Guajira, Magdalena, Putumayo, Nariño and Sucre. Notably, in all cases, after 1965 the estimations of the TFR show a smooth trend.

Figure 2.12: TFR 1958-1971, Colombian macro-regions

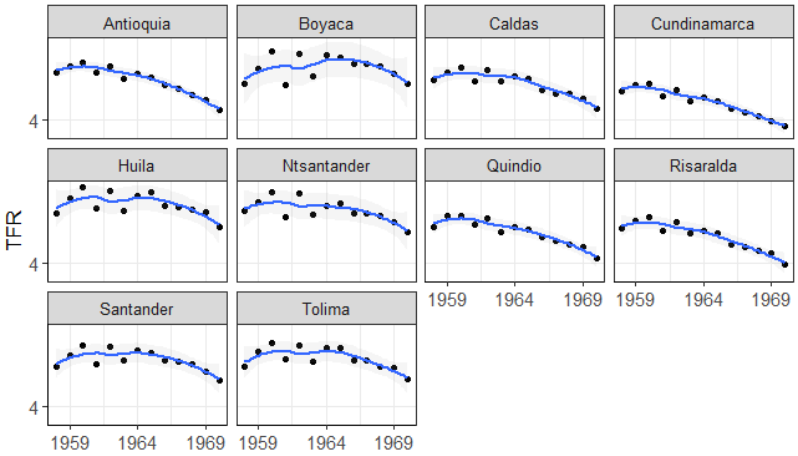
(a) Caribbean



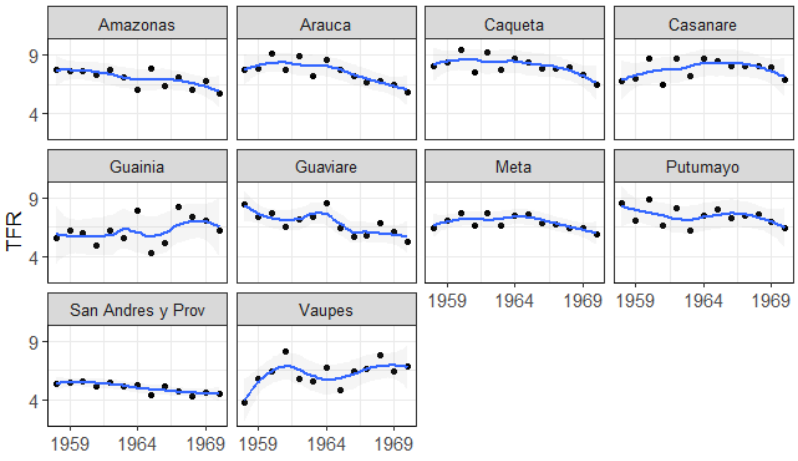
(b) Pacific



(c) Andean



(d) National territories



Note: Calculations using the OCM. Women between 15 to 78 years old and children between 0 to 14 years old. See Section 2.6 for a description of the methodology. The blue line is the loess smooth curve and the black dots represent the TFR estimation for each year. Sources: Authors' calculations based on 1973 full Census data, and life-tables from National Agency of Statistics (DANE).

2.6.4 Fertility in indigenous communities

The census also allows me to calculate fertility measures for women living in indigenous communities. In Colombia, indigenous communities are mostly located in isolated areas where the effects of education, mass media and family planning on fertility are likely to be different from the rest of the country. Most of the existing research on the fertility of indigenous women focuses on contemporary and unrepresentative data and concludes that indigenous populations have higher rates of fertility compared to the non-indigenous population (Peyser & Chackiel, 1993; Arias-Valencia, 2001; McSweeney & Arps, 2005; Arias-Valencia, 2005; Thiede & Gray, 2020; Piñeros-Petersen & Ruiz-Salguero, 1998). I use the 1973 census to estimate fertility rates in indigenous women at different levels of aggregation.

Following the OCM I estimate Total Fertility Rates from 1958-1970 and the results are shown in Fig. 2.13.[32] The main limitation of this source is that I do not observe indigenous women living outside indigenous communities. Indigenous women that migrated, for example, to Bogota cannot be identified. Similarly, age-heaping and under-reporting of children 0 to 1 is problematic and thus I do not include the years 1971 and 1972 in the analysis of the TFR.

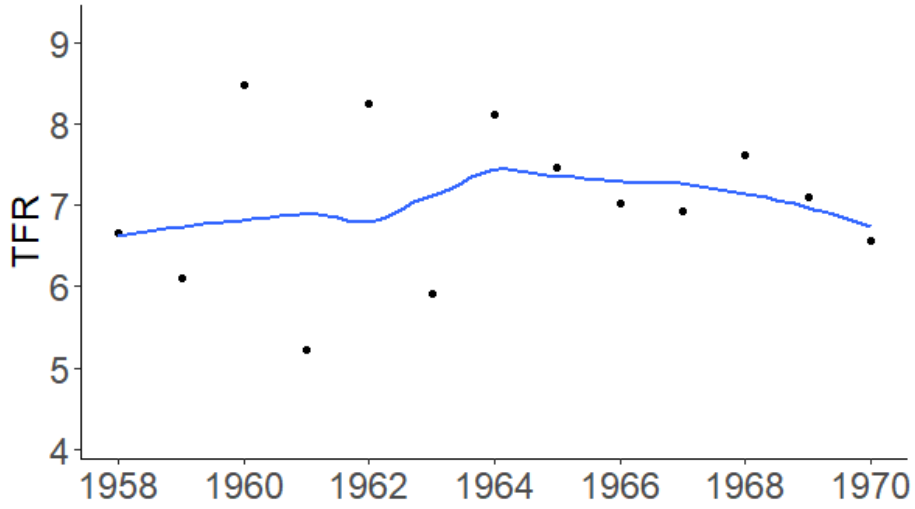
To my knowledge fertility rates for indigenous groups in Colombia have not been estimated for this period. Currently my estimations do not differentiate by ethnic groups but this is a potential avenue for future research. The census reports more than 115 indigenous groups across the five macro-regions, but indigenous communities are concentrated in the Amazonian and Orinoquia with some populations along the Pacific and in La Guajira.

The graph reveals that on average fertility was stable and it fluctuated around 7 children per woman. The estimation shows jagged fluctuations from 1958 to

[32]The mortality adjustments were done at the departmental level using life-tables from 1970.

1964 probably due to age-heaping but this seems to be less concerning after 1964 (i.e. among younger children). These results suggest that indigenous women did not necessarily have higher fertility than the rest of the country as their TFR was relatively low in comparison to several departments and national territories like Casanare.[33] The results reveal that, in contrast to the national trend, indigenous communities did not undergo a significant fertility change during this period. This finding supports the idea that fertility declined in the country as a consequence of a cultural or technological shock that diffused across Colombia but did not reach the more isolated areas where indigenous were located.

Figure 2.13: TFR 1958-1970, Indigenous groups



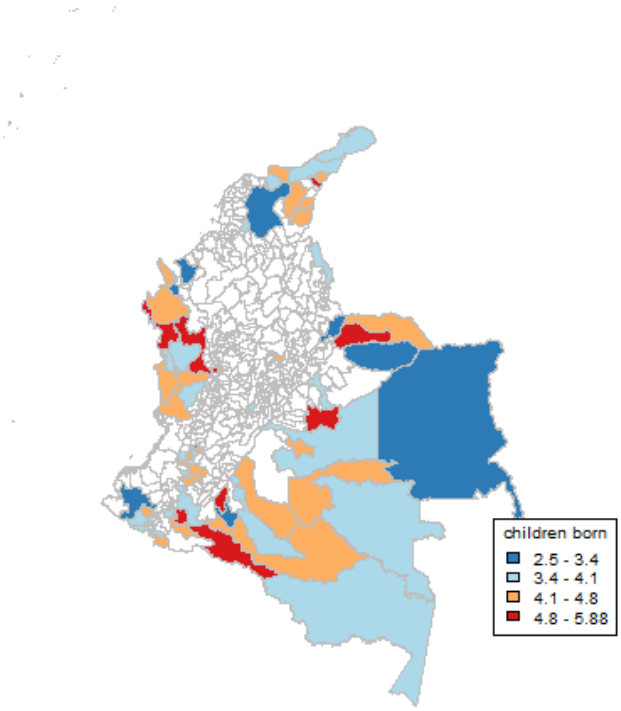
Notes: Note: Calculations using the OCM using. Indigenous women living in indigenous communities are between 15 to 78 years old and their children are between 0 to 14 years old. See Section 2.6 for a description of the methodology. The blue line is the loess smooth curve and the black dots represent the TFR estimation for each year. Sources: Authors' calculations based on 1973 full Census data.

When looking at the number of children ever born in indigenous communities the geographical results in Fig. 2.14 reveal considerable variation in fertility across different indigenous groups. On average, the number of children ever born in 1973 in

[33]Unfortunately I do not observe indigenous women with the 1985 sample of the census from IPUMS. The indigenous information from the 1993 census is still to be processed.

indigenous communities was 4.1, but in some municipalities like Necocli in Antioquia or in the territory of Caqueta the average number of children ever born was as low as 2.5 children. This result is significant by itself as it reveals that indigenous women in Colombia did not necessarily experience higher fertility. Unlike the results for non-indigenous women, there is no clear spatial pattern. Pockets of high fertility appear in the Pacific, the Amazonian and the Orinoquia macro-regions, and low fertility appears also widespread in the territory. Although further investigation of the socioeconomic characteristics of the indigenous groups is required, by 1973 indigenous communities had the worse indicators of development of the country (in terms of education or access to public services). Despite this, fertility levels varied across different indigenous communities.

Figure 2.14: Geographical variation of children ever born in 1973



Notes: The graph shows the average number of children ever born to women living in indigenous communities. The average number of children born was 4.1. Sources: Authors' calculations based on 1973 Census data.

2.7 Conclusions

In this chapter I presented new empirical evidence on the fertility decline of Colombia focusing on the regional perspective. This was key because aggregated fertility measures can hide important regional patterns, especially in a country characterised by stark regional inequalities like Colombia. The findings reveal that geographical and historical differences that characterise the country were also reflected in different fertility patterns before the fertility transition. However, these differences did not influence the start and speed of the fertility transition as the change happened simultaneously across the territory.

Using individual-level data from the complete census of 1973 and 1993 I identified spatial patterns and tested for high and low fertility geographical clusters before and during the fertility transition using Local Indicators of Spatial Association – LISA for women with completed fertility born between 1910-1920 and 1930-1940. The main commonality of the municipalities that form the high-fertility clusters before and during the fertility transition is the neolocal character of the communities. Most of these municipalities, all of them below 1,100 m above sea level, arose after the 1800s as the frontier expanded and migrants establish new societies far from their kinship. On the contrary, the municipalities that were part of the low-fertility clusters have a strong link with the Colonial era and are located in the territory in which both pre-colonial and colonial populations were established in the past. This new empirical evidence suggests that fertility norms before the fertility transition differed in the territory as a consequence of the different historical legacies.

Using the complete registers from the census of 1973 and implementing the Own Child Method, I presented new estimations of fertility rates for Colombia, at the national and the departmental level, including fertility estimations for indigenous communities for the period between 1958 to 1970. These estimations presented a new

picture of the regional fertility decline of the country. At the beginning of the 1960s, fertility declined fast in places where women had traditionally low fertility as well as in places where women had higher fertility. As a consequence, fertility rates almost halve across the country in only 25 years. This is a remarkable finding as it shows that after 1964 fertility declined at a similar rate in regions with different historical backgrounds.

The results at the departmental level confirm that before fertility started declining there was a difference of more than 3 children per woman across departments. In the 1960s fertility started falling simultaneously across the country and by 1970 the fertility transition was on its way almost everywhere, except in indigenous communities. The evidence for indigenous communities shows that fertility rates were stable and fluctuated around 7 children per woman and that in contrast to the national trend, indigenous communities did not undergo a fertility change during this period. These results support the idea that fertility declined in the country as a response to changes in social norms that spread across Colombia but that did not affect indigenous communities that have been more isolated from cultural and technological changes. It is noticeable also that the geographical variation of children born in 1973 reveals considerable variation in fertility across different indigenous groups, additionally, they show that indigenous women did not always have higher fertility than the rest of the country.

The simultaneity and the similar patterns in the fertility decline followed by departments and national territories confirm that this demographic process does not fully reflect improvements in socioeconomic conditions. Moreover, they suggest that other factors, such as the diffusion of social norms, could have played a more prominent role.

The findings of this chapter contribute especially to the debate on the origins

of the fertility decline. Differences in fertility patterns before the fertility transition were explained by geographical and historical factors. However, the fertility transition happened across the whole country suggesting that the decline happened independently of economic development, which was certainly unequal across the country in the mid-1960s.

2.8 What explains the rapid fertility decline?

This chapter shows that fertility started falling simultaneously across all departments and most territories around 1964 and by 1970 the fertility transition was on its way almost everywhere. Even departments like Chocó and Guajira and territories like Putumayo and Guaviare that had a high fertility rate in 1958 (more than 8 children per woman), saw a reduction of at least 2 children per woman in only 12 years. The exceptions are the indigenous communities that show a flat pattern and the departments of Guainia and Vaupes in the Amazon macro-region.

This new evidence confirms that this demographic process does not fully reflect improvements in socioeconomic conditions and suggests that other factors, such as the diffusion of social norms, or sudden technological or cultural shocks, might explain the decline. The next two chapters of this thesis examine two potential mechanisms behind this rapid fertility transition: the introduction of family planning and the increases in education.

Chapter 3 delves into the role of Profamilia in the fertility decline of the country. The chapter estimates the effects on fertility rates of a radio campaign promoting family planning clinics in the late 1960s. The campaign was launched by Profamilia in 1969 to increase visits to the clinics and was broadcast in 13 cities. The 30 seconds radio advertisement gave information about the address of the clinic but did not promote the use of contraceptives nor explain any method. Although the

timing of the campaign coincides with a fall in fertility rates, my results show that the campaign did not affect fertility rates in municipalities with better radio reception or in municipalities close to a Profamilia clinic. These results indicate that knowledge of and availability of contraception did not bring about a faster fertility decline.

Chapter 4 provides new stylised facts of the long-run relationship between females' education and fertility at the national, sub-national and individual-level. I focus on the implementation of educational reforms in Colombia during the so-called liberal period of the 1930s. The findings suggest that the relationship between women's education and fertility is not monotonic. At the individual level, the relationship between education and fertility holds strongly. However, at the national and sub-national level, fertility decline cannot be explained by the direct effects of education as fertility fell continuously in all educational groups since 1965. This suggests that other forces than education were driving the decline.

2.A Appendices

2.A.1 Censuses variables and information on the samples

Table 2.6 gives detailed information on the type of census (full count or sample), variables used for individuals and households, the geographical coverage and the source.

Table 2.6: Censuses information

	1973	1985	1993
Type	Full count	Sample	Full count
Individual variables	Household id, person id, relationship to the head of household, sex, age, marital status, municipality of residence, municipality of birth, department of residence, department of birth, years living in this municipality, schooling, occupation, children ever born, children surviving	Household id, person id, relationship to the head of household, sex, age, marital status, municipality of residence, municipality of birth, department of residence, department of birth, schooling, children ever born, children surviving	Household id, person id, relationship to the head of household, sex, age, marital status, municipality of residence, municipality of birth, department of residence, department of birth, schooling, children ever born, children surviving
Household characteristics	Household id, urban, house construction materials, aqueduct, sewage, electricity, family size, ownership	Household id, urban, house construction materials, aqueduct, sewage, electricity, family size, ownership	Household id, urban, house construction materials, aqueduct, sewage, electricity, family size, ownership
Geographical coverage	All departments and municipalities	All departments and municipalities	Excludes two departments: Guainía and Vaupés
Source	DANE	IPUMS - International	DANE

2.A.2 Linking mothers and children in the household

The set of rules to link a potential mother with a potential child are defined following Sobek and Kennedy (2009). The Integrated Public Use Microdata Series - IPUMS used these rules for their samples (including the 1973 sample of Colombia) to create links between partners, and parents and children. The set of rules and the number of links made following each rule is explained in Appendix Table 2.7. As shown in the first 2 rows, 88% of the links are unambiguous links in which children are matched to women that appear as either partner of the head of the household or as heads of the household.

Table 2.7: Linking rules: Mothers and children in 1973 census.

Rule	Description	Age difference	Number of the links	Percentage of links
Rule 1	If the relationship of the candidate children to the head of household is children and there is no female head, the candidate mother is the woman whose relationship to the head of household is spouse.	10-69	6,868,250	66.6%
Rule 2	If the relationship of the candidate children to the head of household is children and the candidate mother is the head of household.	10-69	2,244,250	21.7%
Rule 3	If the relationship of the candidate's children to the head is grandchildren then the candidate's mother could be a woman whose relationship to the household is child, grandchild or other relative.	15-44	486,057	4.7%
Rule 4	If the relationship of the candidate children to the head is other relative, other non-relative or unknown and the relationship to the head of the household of the candidate mother is other relative	15-45	263,253	2.6%
Rule 5	If the candidate children is the head of the household the candidate mother would be that whose relationship to the head of household is parent.	10-69	207,351	2.0%
Rule 6	If the relationship of the candidate children to the head is other relative, other non-relative or unknown and the relationship to the head of the household of the candidate mother is other non-relative	15-45	109,596	1.1%
Rule 7	If the conditions are similar to rule 1 but the head of the household appears with two or more spouses.	10-54	80,880	0.8%
Rule 8	If the relationship of the candidate children to the head is other relative, other non-relative or unknown and the relationship to the head of the household of the candidate mother is domestic employee	15-45	41,641	0.4%
Rule 9	If there are more than one female heads in the household, take the woman that appears first as the candidate's mother.	10-69	13,650	0.1%
Rule 10	If the relationship of the candidate children to the head is other relative, other non-relative or unknown and the relationship to the head of the household of the candidate mother is grandchild	15-45	3,977	0.0%

Notes: The rules are based on Sobek and Kennedy (2009).

2.A.3 Additional information about the fertility of the clusters

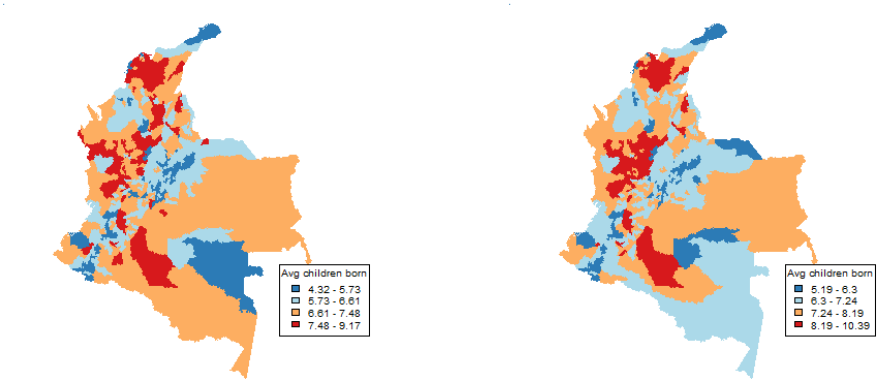
This appendix provides additional information on spatial fertility patterns.

Fig. 2.15 presents the average number of children ever born for women born between 1910-1920 at the municipal level. Fig. 2.15a shows the average number of children ever born looking at all the women in the sample, including those women

without children. Fig. 2.15b present the results of only mothers.

Figure 2.15: Spatial association in completed fertility mothers and all women included

(a) All women (b) Only mothers

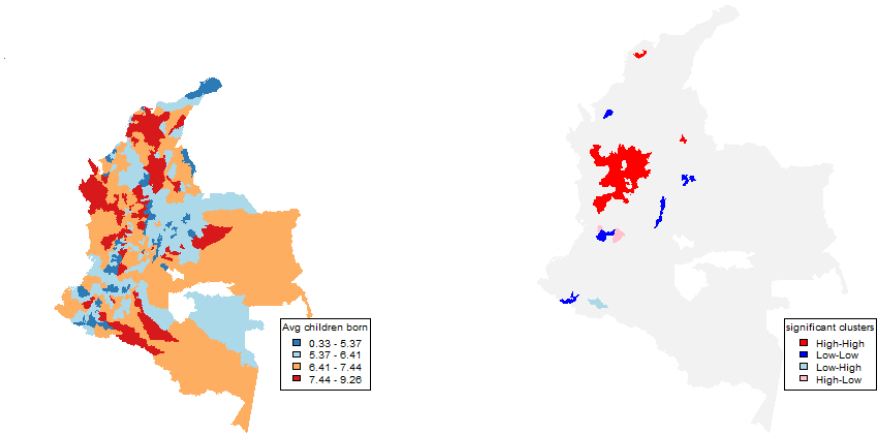


Notes: Fig. 2.15a plots the average number of children by the municipality focusing on all women born between 1910-1920. Fig. 2.15b plots the average number of children by the municipality focusing only on mothers born between 1910-1920. Sources: Authors' calculations based on 1973 Census data.

Fig. 2.16 presents the results of the municipal effects on children ever born when observing non-migrant women. The municipalities of Arauquita and Calamar are those municipalities that appear in white, as they do not report any non-migrant women born between 1910-1920.

Figure 2.16: Spatial association in completed fertility for non-migrant women

(a) Municipality effect (b) Significant spatial association



Notes: Fig. 2.16 plots an average number of children by municipality with only non-migrant women. The municipalities of Arauquita and Calamar are those municipalities that appear in white, as they do not report any non-migrant women born between 1910-1920. The clusters are defined by the effect of the municipality of birth on the number of children ever born to women born between 1910-1920. Fig. 2.6b shows the results of the LISA test. Sources: Authors' calculations based on 1973 Census data.

The following tables list the municipalities that form the geographical clusters. Table 2.8 lists the municipalities that form low fertility clusters before the fertility transition and Table 2.11 shows the municipalities that are part of the pre-transition high fertility clusters. Table 2.10 and Table 2.9 present the municipalities that form low and high fertility clusters during the fertility transition, respectively.

Table 2.8: Municipalities in low clusters 1910-1920 cohort, by department

Departament	Municipalities
Cundinamarca	Facatativá, Zipaquirá, Chía, Mosquera, Madrid, Funza, Cajicá, Cucunubá, Lenguazaque, Sutatausa, Villapinzón, Gachancipá, Tocancipá, Fuquene, Guachetá, Simijaca, Susa, Cota, Tenjo, Ubaté, El Rosal, Bojacá, San Antonio del Tequendama, Zipacón, Cáqueza, Ubaque, Sopó, Tabio, Bogotá
Santander	Charalá, Gambita, Suaita
Boyacá	Cucaita, Sogamoso, Duitama, Combita, Cuitiva, Firavitoba, Iza, Siachoque, Sotaquirá, Toca, Tuta, Viracacha, Garagoa, Guateque, Guayata, Somondoco, Sutatenza, Tenza, Arcabuco, Briceño, Buenavista, Caldas, Coper, Gachantivá, Saboya, Santa Sofía, Samacá, Turmequé, Ventaquemada, Villa de Leyva, Ráquira, Sáchica, San Miguel de Sema, Sutamarchán, Tinjacá, Chita, Mongua, Mongui, Sativasur, Socotá, Topagá, Betéitiva, Busbanzá, Corrales, Floresta, Gámeza, Paz de Río, Socha, Tasco, Boyacá, Ciénega, Jenesano, Nuevo Colón, Ramiriquí, Aquitania, Pesca, Tota, Paipa, Nobsa, Tibasosa, Berbeo, Chinavita, Miraflores, Rondón, San Eduardo, Zetaquirá, La Capilla, Pachavita, Tibaná, Úmbita, El Cocuy, Jericó, La Uvita, Sativa Norte, Susacón, Tutazá
Cauca	Popayán, Villa Rica, Silvia, Totoró, Puerto Tejada, Caldonó, Timbio
Nariño	Nariño, Ipiales, Samaniego, Santacruz, Sapuyes, Providencia, Buesaco, El tablón de Gómez, San Pedro de Cartago, San Lorenzo, Gualmatán, Ospina, Pupiales, Aldana, Cuaspud, Guachucal, Contadero, Iles, Yacuanquer, Imues, Funes, Puerres, Tangua, Consacá, La Florida, Cumbal, Ancuya, Guaitarilla, Mallama, Ricaurte
Valle del Cauca	Cali, Candelaria

Notes: Municipalities in low fertility clusters as defined in Fig. 2.6b

Table 2.9: Municipalities in high clusters 1910-1920 cohort, by department

Departament	Municipalities
Antioquia	Itagui, Envigado, Ríonegro, Angostura, Carolina, Entrerríos, Gómez Plata, Guadalupe, Santa Rosa de Osos, Abejorral, Argelia, Nariño, Sonsón, Fredonia, Valparaiso, Venecia, Caldas, Copacabana, El Carmen de Viboral, La Unión, Santafé de Antioquia, Anzá, Caicedo, Giraldo, Liborina, Olaya, Alejandria, Concepción, Granada, Guatapé, San Rafael, Sanroque, Concepción, La Estrella, Andes, Pueblorrico, La Ceja, Ebéjico, Heliconia, San Jerónimo, Sopetrán, Sabaneta, Marinilla, Caracolí, Maceo, San Carlos, Yolombó, Abriaquí, Murindó, Urrao, Girardota, Betania, Ciudad Bolívar, Barbosa, Peñol, San Vicente, Don Matias, San Pedro, Guarne, Angelópolis, Concordia, Titiribí, Amagá, Jardín, Jericó, Puerto Triunfo, El Santuario, Montebello, Retiro, Salgar, Tarso, Uramita, Caramantá, Támesis, San Francisco, Cisneros, Santo Domingo, Armenia, Belmira, Sabanalarga, San Andrés de Cuerquia, San José de la Montaña
Bolívar	Calamar, Mahates, María la baja, Campo de la Cruz, Santa Lucía, Suán, El Guamo, San Juan Nepomuceno, Manatí, Repelón, Candelaria, Ponedera
Caldas	Marsella, Santa Rosa de Cabal, Guaticá, Quinchiá, Novitá, San José del Palmar, Sipí, Balboa, La Celia, Santuario, Apia, Belén de Umbria, La Virginia, Aranzazu, Filadelfia, Neira, Anserma, Palestina, Chinchiná, Aguadas, Salamina, Belalcazar, San José, Viterbo, Ríosucio, Marmato, Supia, Manzanares, Marquetalia, Marulanda, Pensilvania, La Merced, Pácora
Caquetá	Puerto Rico, Cartagena del Chairá
Chocó	Bagado, El Carmen de Atrató, Lloró, Mistrató
Cesar	El Retén, Ariguaní, Cerro San Antonio, El Piñon, Fundación, Pedraza, Pivijay, Nueva Granada, Remolino, Salamina, Sanzenón, Pijiño del Carmen, Chiboló
Valle del Cauca	Ansermanuevo, Argelia, Toro, La Unión, El Águila, El Cairo, Versalles, Bolívar, El Dovio

Notes: Municipalities in high fertility clusters as defined in Fig. 2.6b

Table 2.10: Municipalities in low clusters 1930-1940 cohort, by department

Departament	Municipalities
Antioquia	Medellín, Itagui, Envigado, La Estrella, Sabaneta
Quindío	Armenia, Salento
Risaralda	Pereira, Dosquebradas, Vill María
Cauca	Puerto Tejada
Cundinamarca	Soacha, Facatativá, Zipaquirá, Chía, Mosquera, Madrid, Funza, Cajicá, Cota, Tenjo, Sibaté, Subachoque, Bojacá, San Antonio del Tequendama, Zipacón, Sopó, Tabio, Bogotá
Nariño	Pasto, Chachaguí, Ipiales, Samaniego, Santacruz, Sapuyés, Providencia, Túquerres, Gualmatán, Ospina, Pupiales, Aldana, Cuaspud, Guachucal, Contadero, Iles, Imues, Yacuanquer, Funes, Puerres, Tangua, Sandona, Belén, La Cruz, Cumbal, Ancuyá, Guaitarilla, Mallama, Ricaurte
Tolima	Flandes
Valle del Ccauca	Cali, Palmira, Yumbo, Jamundí, Candelaria

Notes: Municipalities in low fertility clusters as defined in Fig. 2.8b

Table 2.11: Municipalities in high clusters 1930-1940 cohort, by department

Departament	Municipalities
Antioquia	Angostura, Carolina, Entrerríos, Gómez plata, Guadalupe, Santa Rosa de Osos, Cáceres, Tarazá, Valdivia, El Carmen de Viboral, La Unión, Arboletes, San Juan de Urabá, Santafá de Antioquia, Anza, Caicedo, Giraldo, Liborina, Olaya, Alejandria, Concepción, Granada, Guatapé, San Rafael, Sanroque, Dabeiba, Frontino, Mutatá, Marinilla, Amalfi, Anori, Campamento, Vegachi, Briceño, Toledo, Yarumal, Caracoli, Maceo, San Carlos, Yolombó, Buriticá, Ituango, Peque, Puertotriunfo, San Luis, El Santuario, Salgar, Tarso, Cañasgordas, Uramitá, Cocorná, San Francisco, Armenia, Belmira, Sabanalarga, San Andrés de Cuerquía, San José de la Montaña
Bolívar	Mompós, San Fernando, El Peñón, Talaigua Nuevo, Margarita, Pinillos, Maria la baja, Santa Rosa del sur, Simití, Morales, Ríoviejo, Achí, San Pablo, Barranco de loba, Córdoba, Zambrano, Coloso, Chalán, Ovejas
Caquetá	Cartagena del Chairá, Puerto Rico, San Vicente del Caguán, Curillo, Solano, Valparaíso, La Montañita, Milán, Morelia, Albania, Belén de los Andaquies, San José del Fragua
Cesar	Abrego, Cachirá, La Playa, Chiriguaná, El Paso, La Jagua de Ibirico, Astrea, Chimichagua, Pailitas, Pelaya, Tamalameque, El Tarra, Hacarí, Sancalixto, Teorama, Bucarasica, Salazar, Santiago, Sardinata, Villacaro, Curumani, Convención, El Carmen, Guamal, San Sebastián de Buenavista
Córdoba	Canalete, Los Córdobas
Huila	Elias, Oporapá, Tarquí, Acevedo, La Argentina, Saladoblanco
Nariño	El Charco, Latola, Santa Bárbara

Notes: Municipalities in high fertility clusters as defined in Fig. 2.8b

3

Waves of change? Radio announcements and fertility decline

3.1 Introduction

Currently, there is no consensus regarding the role of contraception and family planning programs in the fertility decline of developing economies.[1] Several scholars argue that family planning can have substantial effects on fertility and suggest that knowledge and availability of contraceptive methods are one of the most fundamental barriers to fertility decline (Bongaarts et al., 1990). On the contrary, others suggest that once changes in the demand for children are accounted for, the effect of contraceptives on fertility is small (Miller & Babiarz, 2016). For the case of Colombia, Miller (2010) estimates that the foundation of a Profamilia clinic in a city accounted for around 6% to 7% of the decline by looking at the geographical spread of the organisation from 1965 to the beginning of the 1970s.

This chapter investigates if a national radio campaign promoting family planning clinics in the late 1960s in Colombia had an effect on fertility across the country. This initiative by Profamilia started in 1969 and introduced publicly the concept of family planning. The radio campaign increased the availability of family planning information, especially in rural areas, where radio was the primary source of information. However, the campaign did little in promoting new social roles for women or in increasing the educational level of the population and it was mostly used as an attempt to make the population both aware of and interested in the programs and clinics.

The use of new media and information technologies has the potential to spread information about family planning and at the same time it could shape listeners' attitudes. It can promote family planning and can alter the behaviour of people,

[1] It is likely that contraception played an important role during the historical fertility transition in Western countries, although mainly by the implementation of withdrawal and abstinence given the high costs of condoms or the diaphragm (Guinnane, 2011). For the US case, scholars have argued that new contraceptive methods like the pill did not play a major role in the sharp drop in fertility after the 1960s (Bailey, 1973; Angrist & Evans, 1998).

changing the demand for children. Media exposure can disseminate and increase the availability of information about family planning while persuading listeners, and at the same time it can shape new social roles, change social norms and raise the educational level of the population.[2]

As shown in Chapter 2, the fertility transition in Colombia started in the mid-1960s. The change was fast and widespread and fertility declined almost simultaneously in regions with traditionally high fertility levels as well as in regions with lower fertility rates. Not only in Colombia, but across the world fertility rates started decreasing faster than ever. The timing of the decline coincided with the introduction and popularisation of modern contraceptive methods and in particular with the release of birth control pills in the US. Although by the mid-1960s the pill had spread quickly across the world, especially in urban centres, the practices of fertility control were not common in rural areas. Also, the knowledge and use of contraception methods varied widely across the world. However, after 1965, fertility declined rapidly both in rural and urban areas in Colombia and the reasons behind such a rapid decline have not been yet addressed.

The radio campaign started in 1969 with 30-seconds ads that announced the availability of family planning services and locations of Profamilia clinics. The idea was to spread the idea of family planning and increase the number of visitors to the clinics. The ads did not describe any contraceptive method nor intended to persuade the listeners of the advantages of limiting fertility. Early evaluations of the radio campaign show that it reached the urban target population and accelerated information diffusion about fertility control and increased awareness of contraceptive methods (Bailey, 1973; Stycos & Avery, 1975; Bailey & Cabrera, 1981). Overall, the

[2]Recent literature has shown that the use of new media and information technologies has the potential to affect socioeconomic outcomes such as social capital (Olken, 2009), political outcomes (Wang, 2021), inter-group animosity (Adena, Enikolopov, Petrova, Santarosa & Zhuravskaya, 2015), and fertility decisions (La Ferrara, Chong & Duryea, 2012).

evaluations show an immediate increase in the number of visitors to the clinics during the period of the radio campaign, especially in large cities, but some suggest that this increase was at the expense of visitors that would have attended the clinics in the future (Stycos & Avery, 1975). Until now, the effect of this campaign on the rapid fertility decline of Colombia has not been measured.

This chapter examines whether Profamilia's radio campaign spurred fertility decline by raising awareness of the availability of contraceptive methods and legitimising the concept of family planning across the country and particularly, in rural areas that were close to cities with family planning clinics.

To evaluate the effects of exposure to radio campaigns, this chapter exploits exogenous variation in radio signal strength that results from topographic factors. I use individual-level data from the full count census of 1973 and collected data on the location and dates of the establishment of Profamilia's clinics. Given that I do not have exact data on the number of radio listeners or variation in access to radio at the individual level, I gathered information on the content, coverage, and timing of Profamilia radio programmes, and using the Irregular Terrain Model (Hufford, 2002) I predict the radio signal strength of Radio Reloj, a traditional radio station in the country that was located in the cities were Profamilia launched its radio campaign, and use the predicted signal strength to estimate a lower bound of the intention to treat effect.

The main challenge to estimate the effects of the radio campaign on fertility is that it is possible that the location of the stations was not randomly allocated in the country, and therefore it could be correlated with other characteristics that could influence fertility. Although the configuration of the broadcasting stations was completed before the introduction of the campaign and previous research has shown that the spread of Profamilia was independent of fertility, I follow different empirical

strategies to minimise this concern. First, I adopt a strategy developed by Olken (2009) that has been used extensively in the literature (DellaVigna & La Ferrara, 2015; Wang, 2021). I regress fertility rates on the predicted radio signal strength while controlling for the hypothetical signal strength where there are no orographic obstacles such as mountains. Following this strategy, the identification of the effect should come from the residual variation of the signal strength as a result of idiosyncratic topographic factors along the signal transmission route. Second, I employ a difference-in-differences strategy to compare fertility rates before and after the start of the radio campaign. To avoid the potential correlation between the location of the clinics and radio transmitters and other characteristics that could influence fertility at the municipal level I exclude cities with transmitters in my analysis. The results show that there was no effect of the radio campaign on fertility rates across the country. Finally, I restrict the analysis to municipalities that were close to a Profamilia clinic to observe if the radio campaign affected fertility in these municipalities as they benefited from having access to the clinics. The results confirm that the Profamilia radio campaign had a limited effect on fertility rates even in municipalities located close to cities with a family planning clinic.

This research contributes to several debates in the literature. The first one is related to the effects of family planning programs on fertility decline. Although global funding for family planning tripled during the 1970s and early 1980s the effect of these programs on fertility is controversial (Miller & Babiarz, 2016). Some scholars argued that the large scale initiative to implement family planning programs across the world was successful in reducing fertility and population growth on a global scale.[3] However, the estimations at the country level range from no effect in the case of Indonesia in the 1970s and 1980s (Pitt, Rosenzweig & Gibbons, 1993) to 27% in the case of China

[3]For example, Bongaarts et al. (1990) estimate that without family planning programmes the population in Asia, Latin America, and Africa would have been 412 millions bigger by the early 1990s.

with its "Longer, Later, Fewer" campaign (Babiarz, Ma, Miller & Song, 2020). For the case of Colombia, my research confirms that access and knowledge of contraceptive methods had a limited effect on the fertility decline of the country.

This research is also related to the literature on the effects of media and information technologies on fertility. Mass media has the potential to alter the behaviour of people by providing new information but also by shaping the self-identity of people, also known as ideational change (Williams & Singh, 1976; Barber & Axinn, 2004; Spolaore & Wacziarg, 2022).[4] For example, the literature finds that there is a significant association between mass media and fertility (Williams & Singh, 1976; Westoff & Rodriguez, 1995; Parr, 2002; Beach & Hanlon, 2019). At the cross country-level access to television, radio and newspapers are negatively related to fertility rates. Several reasons for these relationships have been suggested in the literature. Potentially, media can directly disseminate information about family planning, as in the case of my study or it can shape social roles, by presenting alternative lifestyles, examples of different aspirations or by raising the general level of education of the population. The paper by La Ferrara et al. (2012) on the case of Brazil is a clear example of this literature.[5] The study shows how the spread of soap operas affected the Brazilian fertility transition and how *telenovelas* spread a new idea of family and in particular raised women's economic aspirations.[6] Profamilia's radio campaign had a limited role in disseminating new social roles or in raising the economic aspirations of the population and it was used mostly as an attempt to make the population both aware and interested in the programs and clinics. Given that the campaign did little in terms of promoting new social roles, a process of imitation, as in the case of soap

[4]See also DellaVigna and La Ferrara (2015) for a recent survey on the impact of exposure to the media.

[5]See also Barber and Axinn (2004) on Nepal or Kearney and Levine (2015) for the US case.

[6]*telenovelas*, were developed in Mexico in the 1970s by Miguel Sabido, a producer in a Televisa, a Mexican multimedia mass media company. This type of tv drama spread quickly across the continent.

operas in Brazil, was difficult to accomplish.

This finding is also related to the seminal paper on Nazi radio and its effects on anti-Semitism by Adena et al. (2015) or the more recent papers on the effects of radio on political outcomes during the 1930s and 1960s in the US by Wang (2021) which underline the importance of persuasion in communication.[7] Unlike these studies, this chapter focuses on a campaign of information that had limited power of persuasion. The campaign was focused on giving information on the location of the clinics and legitimising the idea of family planning and the announcements did not intend to persuade couples of the advantages of smaller families.

Finally, using both exogenous variations of signal strength and a difference in difference strategy, my research provides robust evidence of the limited influence of family planning promotion through radio on fertility behaviour. Since the 1850s family planning advocates have used mass media outlets to inform and motivate people on the methods and advantages of regulating fertility (Parry, 2013; Beach & Hanlon, 2019).[8] But with the rise of new media and information technologies in the 1950s the use of radio and TVs became popular to promote modern contraceptives given their wide coverage and potential effectiveness. The evidence of such effectiveness is mixed (Udry, Clark, Chase & Levy, 1972). Additionally, the interpretation of the evidence as causal could be limited as identification of exposure is not always clear.[9]

[7]See Dellavigna and Gentzkow (2010) for a recent survey in the effects and drivers of persuasive communication.

[8]For example, the dissemination of the Bradlaugh-Besant trial after 1877 in the United Kingdom or the pioneer pamphlet of 1914 called *Family limitations* that circulated in the US published by Margaret Sanger and was followed by the Comstock Laws.

[9]See for example the discussion on Jaeger, Joyce and Kaestner (2020) and the study for Ghana by Parr (2002).

3.2 A brief history of the radio

As one of the biggest revolutions, the radio arrived in the country in 1929 and expanded through the territory during the 1930s. The first official (from the government) radio station, the HJN, was founded in Bogotá on the 5th of September 1929, while the first commercial radio station, *"La Voz de Barranquilla"*, was founded in Barranquilla on the 8th of December of the same year. The central goal of the radio was to consolidate a national project and to overcome the geographical barriers to unify the country. Public radio stations were mainly used for educational purposes and to connect the country while commercial radio was mainly used for advertisements (Blanco Sánchez, 2018).

By 1945 the radio had already established itself as the main means of mass communication in Colombia and with the introduction of the transistor radio in the mid-1950s, the radio became the most popular electronic communication device of the 1950s and incorporated the rural area to the national audience. In 1947 the foundation of *Radio Sutatenza*, the first Catholic radio station in Latin America, started an educational campaign through radio to alleviate the high levels of illiteracy in the rural and isolated areas (Roldán, 2014).

After 1957 almost all the radio stations had modernised their equipment and increased their power from 1 or 5 kilowatts to 10 or 20 kilowatts, which improved the coverage range of the radio. By the mid-1970s the radio was regarded as the method par excellence for reaching the lower-income population with around 20 million listeners across the country and 340 radio stations located in 109 municipalities (Ferreira & Straubhaar, 1988). The report of the 1973 census shows that 46% of indigenous families had a radio in the house.[10]

The commercial radio station *Radio Reloj* was first established in Bogota on

[10]Unfortunately this question is absent from the private household form.

the 18th of March of 1949, and expanded to other cities in the 1950s so by the late 1960s it was present in more than 15 cities. The station broadcast news and music throughout the day and was recognised for its particular way of telling the time.

3.3 The *debut* of family planning

Catching up with Europe, like other Latin American countries, in the 1960s Colombia experienced a rapid fertility transition, and in only 25 years the number of children per woman fell from 6.8 to 3 (López Toro, 1968). Despite important regional disparities in development, the fertility decline in the country was not only rapid but also widespread. Both rural and urban women decreased their number of children and the reduction was evident in women of all ages. What explains such a rapid and widespread decline in fertility? There is still an open debate in the literature regarding the mechanisms that can account for the fertility decline but the increase in the knowledge, availability and use of contraceptive methods after the mid-1960s has drawn special attention as the fertility transition coincided with the approval of the pill by the FDA in the US in 1960 and its quick dissemination around the world.

In Colombia, the dissemination of modern contraceptive methods kicked off when the Colombian Association for Family Welfare (Profamilia) was founded in Bogota in 1965. As a result, by the early 1970s, the majority of women knew of at least one contraceptive method and approved family planning, and around 40% of urban women were practising contraception (Prada & Ojeda, 1987; Bailey, 1973). Although the use of modern contraceptive methods increased, the role that Profamilia played in the overall fertility decline was limited. Miller (2010) estimates that the introduction of Profamilia clinics in urban centres explains only between 6% and 7% of the fertility decline between 1964-1993. It is possible that family planning had a small effect on the fertility of urban women because by the early 1960s fertility in

urban areas was already being controlled, possibly through more traditional methods and abortion. But the effects of Profamilia clinics on the fertility of women living in areas outside the city, which were on average more rural and less educated and where access to contraception was more limited, are still unknown.

As reported in Simmons and Cardona (1974), in 1969 63% of women living in Bogota had ever used a contraceptive method, while only 19% of rural women had. Similarly, abortion was commonly practised in big cities during the 1960s (Mendoza-Hoyos, 1968). According to Mendoza, more than 60% of abortions happened in women with seven or more births, suggesting then that abortion was a common practice to limit the number of children. Interestingly, abortion was present mainly in main cities where the induced abortion rate per pregnant woman was around 20%, while only 8% of women living in rural areas practised induced abortion (Requena, 1968).

As shown in Table 3.1, despite the difference in the practice of contraception and abortion, Simmons and Cardona (1974) show that in 1969 women living both in rural and urban areas had similarly favourable attitudes to family planning. This could indicate that the adoption of contraception in rural areas was limited by the availability of contraceptive methods and not so much by awareness or knowledge and that once access to contraception was ensured, the use of contraceptive methods should have increased for rural women. However, an increase in the utilisation of modern contraceptive methods after the establishment and promotion of family clinics does not necessarily imply an effect on fertility if couples were already effectively controlling their fertility.

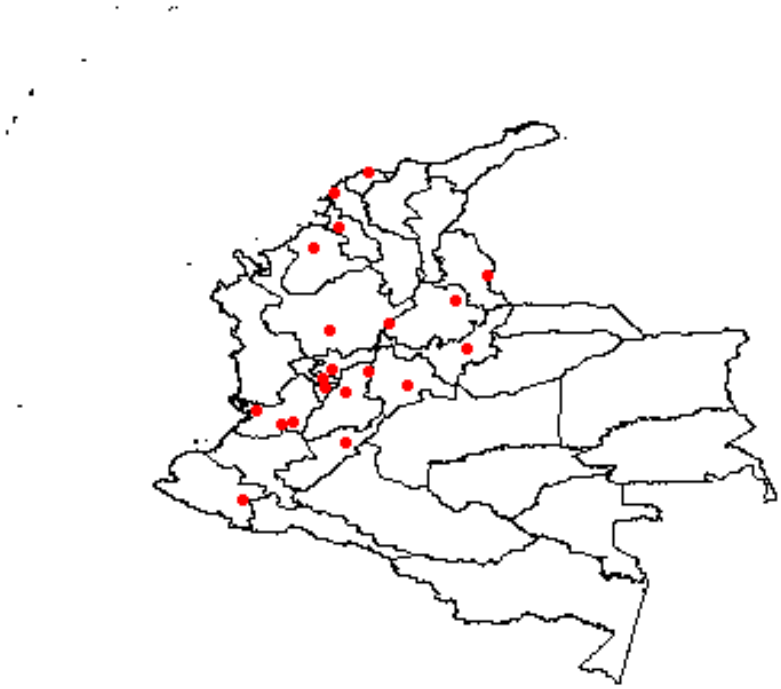
Table 3.1: Favourability to family planning, 1969

	Rural	Cartagena	Medellin	Bogota
Favourable to family planning	62	69	58	76
Have ever used a contraceptive method	19	48	52	63

Notes: Percent of respondents by place of residence in 1969. The table shows the difference in favourability and use between regions in Colombia, and small rural areas. Source: Simmons and Cardona (1974).

By 1969 Profamilia had established 17 clinics in urban centres covering most regions of the country, except for the National Territories (Fig. 3.1). In the same year, the organisation launched a radio campaign in 13 of the 17 cities in which they had a clinic. The main objective of the radio campaign was to spread the idea of family planning in Colombia and "to change the attitudes of Colombians toward the rational control of the size of their families" (Stycos and Avery, 151:1975). The campaign announced the availability of family planning services and locations of Profamilia clinics but it did not describe any contraceptive method. The announcements were broadcast several times a day between 8 am and 6 pm along with other commercial spots and were between 15 to 30 seconds long. Since the start of the campaign the ads were usually broadcast during the second six months of each year but due to financing constraints, there were occasional suspensions of broadcasting for weeks. The 1971 campaign was broadcast over 51 radio stations and every station aired one spot each half-hour, which means that about 900 announcements each day were aired (Stycos & Avery, 1975).

Figure 3.1: Profamilia clinics in 1969



Notes: The map shows the geographical location of Profamilia clinics by 1969. Sources: Based on Miller

There were different types of announcements but the following examples from Bailey (1973) show that the main message of the radio spots was to make the population aware of the ideas of family planning and fertility limitation while giving the exact location of the family planning clinics. These announcements introduced the concept of family planning and fertility control, raised awareness of the availability and by being so frequently on the radio, legitimated the concept of family planning.

> ***Announcer 1***: *Every child needs special attention. Therefore have only the number of children you can take care of.*
> ***Woman 1***: *Talk with your husband and go to Profamilia.*
> ***Woman 2***: *[Gives the address of a clinic]*

> ***Announcer 1***: *Do you know what family planning is?*

***A woman**: Family planning means to have the number of children one wants and to have them when they are wanted*
***Announcer 1:** [Gives the address of a clinic]*

The idea of this campaign was to promote the clinics so that couples that did not want any more children could have information about where to go to get access to family planning services. Given that in the 1970s radio was the best method for reaching lower-income and rural populations the campaign could have been successful in making the population both aware and interested in family planning programs, but as the campaign did little in promoting new social roles, increasing the educational level of women, persuading women or couples of the advantages of smaller families or raising the economic aspirations of the radio listeners, the impact of the campaign in the fertility decline of the country could have been limited. Additionally, it seems plausible that the information about family planning and fertility control was not new to radio listeners. As previously discussed, urban women were already limiting their fertility using other methods such as abortion and rural women were not only aware but also in favour of family planning.

In the eyes of Profamilia, the campaign was successful as the number of users increased soon after the radio campaign started and most importantly, it did not face any opposition from the public or the church, which demonstrated the broad acceptance of family planning in the country. Early evaluations of the radio campaign confirmed that it reached the urban target population and that it accelerated information diffusion about fertility control. Given that broadcasting was not cheap and Profamilia was a private initiative, financed particularly by USAID and Planned Parenthood Federation, most of these evaluations were concerned with the cost-effectiveness of the campaign. The evaluations estimated that the cost per new client was between $8 and $17 US dollars (Bailey, 1973; Stycos & Avery, 1975).[11]

[11]The cost per client of the radio campaign was considerably low in comparison to the estimations

In general, the evaluations find that the radio campaign added a substantial number of visitors to the clinics. Bailey (1973) showed that the number of visitors increased in the urban clinics immediately after the beginning of the campaign and that radio was the second most important source, after friends, neighbours and relatives, of information about the existence of the service. Bailey and Cabrera (1981) calculate that between 1971 and 1972 more than 8,000 new visitors went to a clinic because of the radio campaign, although it is possible that this increase was at the expense of visitors that would have attended the clinics in the future. Also, the major impact of the campaign was in two big cities, Bogota and Medellin, which accounted for almost 60% of the campaign's impact on the number of new visitors. Stycos and Avery (1975) argue that although the radio campaigns likely accelerated the process of information diffusion about fertility control, the campaign was less effective in smaller cities which suggest that informal channels of communication were sufficient to diffuse the information about the location of the clinics. On the whole, the evaluations seem to suggest that although the number of visitors increases in some cities, the overall effect of the campaign could be limited.

3.4 Data

To measure the effects of the national radio campaign on fertility I have assembled a unique data set using several sources. I collect data on the location and dates of the establishment of Profamilia's clinics and radio programmes. I gather data on transmitters' location, frequency, and power for the complete broadcasting network in Colombia and calculate radio signal strength in the country during the 1970s to measure exposure to the radio campaign. Using the full count census of 1973 and life tables I calculate several measures of fertility to address the potential heterogeneous

of the 1971 campaign in the US by Udry et al. (1972) that estimated an average cost per additional new client between $75 and $5,000 US dollars.

effects of the campaign.

3.4.1 Exposure to the radio campaign

Using data from Miller (2010) I compile the location and dates of establishment of Profamilia clinics from 1965 to 1970. From Bailey (1973) and Stycos and Avery (1975) I collected the cities where the radio campaign started in 1969.

Although by the 1960s radio was very popular across the country the information on how many people had a radio at home, listened to the radio or the radio announcements is not available.[12] To overcome this, I exploit continuous variation in signal strength that results from topographic factors and proxy radio exposure in each municipality by a measure of its signal strength. Using the first National directory of broadcasting stations of 1976 published by the National Office of Statistics, I collect information on the radio stations that had transmitters in all cities that broadcast the radio campaigns. Although this information comes from 1976, according to a report from DANE (1978), by 1957 most of the radio stations had increased their power from 1, 2 or 5 kilowatts to 10 or 20 kilowatts and those local radio stations that were unable to improve their coverage were mostly absorbed by bigger radio broadcasters. Therefore, by 1964 the configuration of the broadcasting stations was almost invariant except for some new independent radio stations that emerged after 1964 but these stations were not included in the data I collected.[13]

Given that I don't have information on the exact radio stations used by Profamilia to broadcast the radio announcements, I use data by city on transmitters' location, frequency, and power from one of the more long-established radio stations

[12]Table 3.7 presents statistics compile by Stycos and Avery (1975) on exposure of new visitors to radio. The table shows that on average 89% of women that went to a Profamilia clinic during July and August of 1971 listen to the radio, 40% of these women listen to the radio more than 36 hours per week, and 83% heard one of Porfamilia's spots.

[13]These radio stations are Radio Cadena Independiente, Coral, Super, Cadena Lider de Colombia and Radio Sistema Colibri.

in the country, *Radio Reloj*. It seems reasonable to use a popular radio station such as Radio Reloj to measure the effect of the radio announcements on fertility, given that as Stycos and Avery (1975) mention, Profamilia did achieve coverage of the most popular radio stations in the cities.

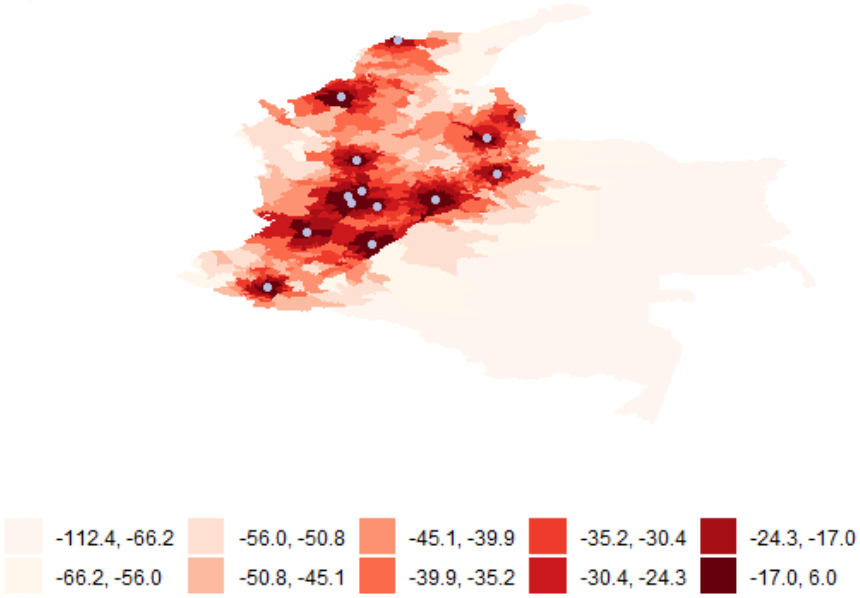
To exploit the exogenous variation in radio signal strength that results from topographic factors, I use the Irregular Terrain Model (ITM) developed by Hufford (2002).[14] Using a professional radio software, the model combines data on the location and characteristics of radio transmitters and the orographic characteristics of the country using a 1 km grid elevation data and information on surface refractivity, radio climate and conductivity of the ground to predict radio signal strength isolating the part of the radio reception that is lost due to the topographic obstacles. For each municipality-radio station pair, the ITM calculates the predicted signal power a receiver would get including the effects of topography and distance to the transmitter.[15] I use the maximum predicted signal strength in each municipality across all transmitters as the predicted signal strength in that municipality. Additionally, the model also calculates the hypothetical signal strength in free space, that is, the signal strength that a municipality would receive if there was no interference from geography or climate. The Fig. 3.2a shows the maximum predicted signal strength of Profamilia transmitters to all municipalities in the country and Fig. 3.2b shows the hypothetical signal strength in free space and Fig. 3.6 shows the distribution of the predicted signal strength.

[14]I would like to thank the generosity of Benjamin Olken who kindly shared with me the ITM software.

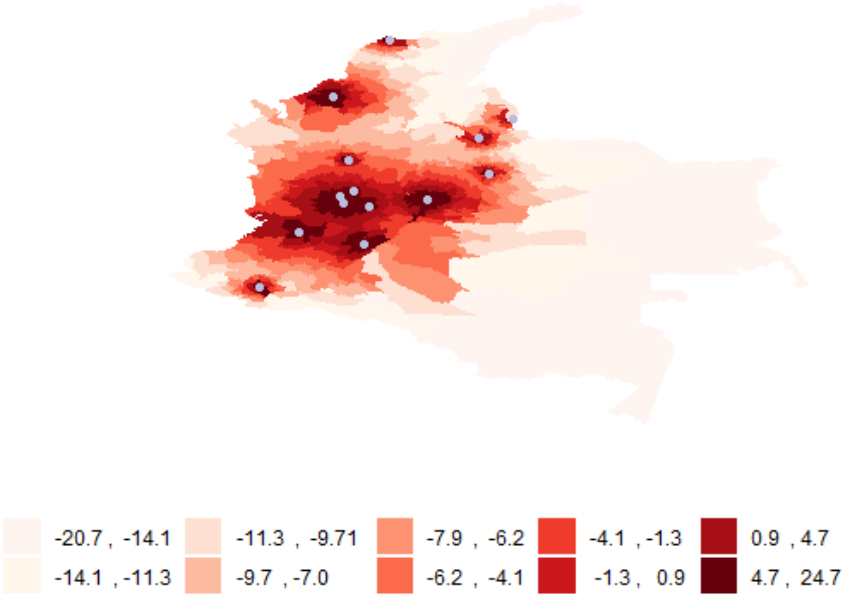
[15]Following Olken (2009), Wang (2021) and others I use the centroid of each municipality as the receiver location.

Figure 3.2: Location and signal strength of Profamilia transmitters

(a) predicted signal strength



(b) free space signal strength



Notes: The maps show the maximum predicted signal strength and the free space signal strength for each municipality for the radio station *Radio Reloj*. The predicted and the hypothetical signal strength are computed using the Irregular Terrain Model (ITM) and measured in decibel-milliwatts (dBm). The grey dots are the transmitter location. Sources: Authors' calculations based on Miller (2010), Bailey (1973), Stycos and Avery (1975) and the National directory of broadcasting (1976).

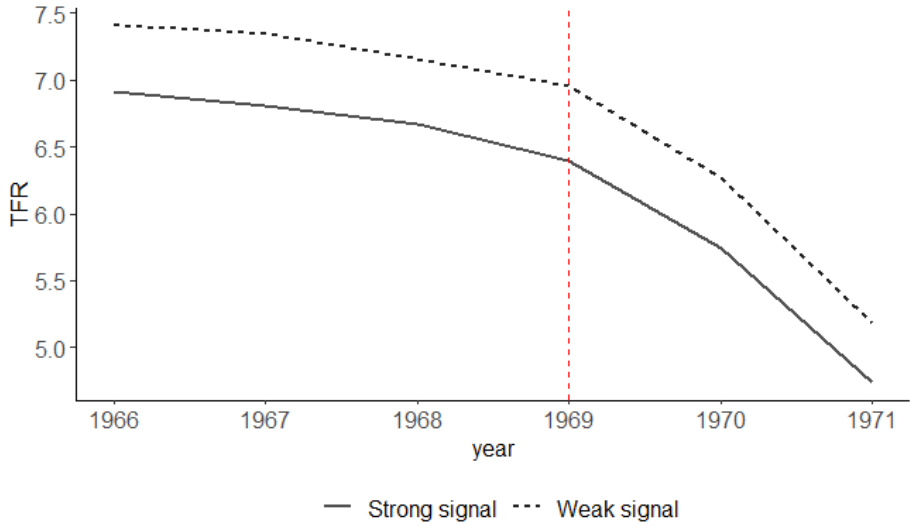
3.4.2 Fertility measures

I use data from the full count census of 1973 and life tables to observe the short-term effects of the radio campaign on fertility. The first outcome of interest is a general fertility rate calculated as the births per women aged 15 to 48. I divide the number of children between 0 to 3 years old by the number of women that are between 15 to 48 years old in each municipality.[16] I use this measure instead of the traditional Crude Birth Rate to account for differences in sex ratios and in age distribution which could be more problematic at the municipal level. The birth rate in the country was on average 611.6 live births per 1,000 women aged 15 to 48 and the distribution is shown in Appendix Fig. 3.8.

To observe the evolution of fertility over time, I compute the Total Fertility Rate (TFR) at the municipal level following the methodology discussed in the previous chapter. I present the TFR from 1966 to 1971 and I use the 1970 life-tables from CEPAL at the departmental level to do mortality adjustments. I focus on children that are between 0 to 7 to minimise the bias in estimates of the total fertility rate (Reid et al., 2020; Timæus, 2021). These children are more likely to live at home with their mothers and therefore are more likely to be matched to the mother minimising measurement error and providing a more robust estimation of the effect. Fig. 3.3 shows the Total Fertility Rate for 1966 to 1971 divided by places with a signal strength above the median signal strength of the sample and places with a signal strength below the median. The figure confirms that fertility declines fast and across different municipalities and it shows that municipalities with a signal strength above the mean had on average lower fertility in comparison to the municipalities with weaker signal strength. Noticeably, fertility declines faster after 1969 in both groups but the trends seem to suggest that fertility declines faster in places with weak signal strength.

[16]The general fertility rate can be calculated using children born to women aged 15 to 45. However, I include children aged 0 to 3 to account for the potential misreporting of ages at ages 0 and 1. Therefore, I also include women aged 45 to 48 to allow for births happening in the previous 3 years.

Figure 3.3: Total Fertility Rate, 1966-1971



Note: The figure shows the Total Fertility Rate (TFR) by the strength of the radio signal received. Calculations using the OCM and based on women between 15 and 64 years old and children between 0 and 7 years old. The estimations of the TFR are adjusted by mortality rates using life-tables from 1970. The dotted line shows the TFR for municipalities with a predicted signal strength below the median signal strength of the sample. The red line indicates the year in which the radio campaign was introduced. Sources: Authors' calculations based on 1973 full Census data.

Additionally, I compute Age-Specific Fertility Rates for the same period and at the municipal level to observe potential heterogeneous effects by age. In particular, I calculate ASFR for ages 15 to 19, 30 to 34 and 40 to 44. These age groups allow me to observe differences in fertility at different points in the fertility life of women. The first group shows the effect of the radio campaign on teen pregnancy and starting. The second group displays the effect on fertility for women that already started their fertility life, while the third group shows the effect of the radio campaign on stopping. If the radio announcements were indeed disseminating new information and new social norms, the potential heterogeneous effect could also come from differences in the adoption of contraception according to differences in age. For example, it is likely that younger women that grew up with constant information about family planning adopt contraception more easily in comparison with older women.

3.5 Empirical strategy

The baseline of my empirical strategy is to measure the effects of exposure to Profamilia's radio campaign on fertility during the onset of the fertility transition.

It would be ideal to measure the effect at the individual-level, however, I do not have exact data on the number of radio listeners or variation in access to radio at this level of information. Using the Irregular Terrain Model (Hufford, 2002), I predict the radio signal strength of Radio Reloj, which was located in the cities where Profamilia launched its radio campaign. With the predicted signal strength as a proxy, I estimate a lower bound of the intention to treat the effect.

The main challenge of identification is that the location of the radio stations and the Profamilia clinics may not be randomly distributed in the country and that these locations could be correlated with other characteristics that influence fertility at the municipal level. As for the location of Profamilia clinics, Miller (2010) shows that the process of expansion of Profamilia was mostly arbitrary as it did not follow any clear geographical pattern (See Fig. 3.1). Furthermore, Miller (2010) demonstrates that programme placement was not correlated with pre-existing trends in socio-economic conditions of the municipalities.

Regarding the placement of the radio stations, the configuration of the broadcasting stations was completed in 1964, 5 years before the radio campaign started. However, to minimise further concerns of identification, I follow different empirical strategies.

First, I follow the empirical strategy pioneered by Olken (2009) to exploit the exogenous variation of the signal strength that results from topographic factors (DellaVigna & La Ferrara, 2015; Wang, 2021). To do so, the first model regresses fertility outcomes on the predicted signal strength calculated at the municipal level.

To isolate the effects of topography on the transmission patterns, the model includes a control for the hypothetical signal strength in free space, that is, the signal strength without obstacles such as mountains. In this model, the identification of the effect comes from the residual variation of the signal strength as a result of idiosyncratic topographic factors along the signal transmission route. The specification is as shown in Eq. (3.1):

$$Y_m = \alpha + \beta * \text{signal}_m + \gamma * \text{Free}_m + \mathbf{X}_c\lambda + \delta_d + \epsilon_m \tag{3.1}$$

Where Y_m is births per women aged 15 to 48, $Signal_m$ is the actual signal strength measured in decibel milliwatts (dBm), $Free_m$ is the hypothetical signal strength without obstacles and δ_d are departmental fixed effects to control for any differences across regions. $\mathbf{X}_c$ is a vector of municipal geographical and historical controls and urbanisation and population controls. Other unobserved factors are captured with the random error term ϵ_m. The standard errors are bootstrapped and clustered at the departmental level. As a robustness check, I estimate the same model but instead of using a continuous measure of the signal strength, I measure signal strength using dummy variables corresponding to different percentiles of the signal and the results are shown in Appendix Fig. 3.11.

As an alternative strategy, I employ a Difference-in-Differences design to compare fertility rates before and after the start of the radio campaign. To address potential concerns of endogeneity in the treatment, I exclude cities with transmitters in my analysis. In this model, the identification of the effect comes from comparing changes in fertility rates in municipalities that received a strong radio signal (treated) with the changes in fertility rates of the municipalities that received a weak radio signal. The difference-in-differences (DD) setup is indicated in equation Eq. (3.2).

$$Y_{m,t} = \alpha_m + \theta_t + \beta \text{strong signal}_m \times post + \epsilon_{m,t} \tag{3.2}$$

Where $Y_{m,t}$ is some measure of fertility in municipality m in year t. For example, $Y_{m,t}$ is the Total Fertility Rate from 1966 to 1971 or the Age-Specific Fertility Rate for different age groups. The main coefficient of interest is the interaction term between the post-treatment indicator ($post_t$) and treatment status strong signal_m, where treated-status means that the signal strength of the municipality is above the median signal strength in the case of a binary treatment. Although the campaign started in 1969, it is only after around 9 months that we expect to see an effect, given that fertility takes time to respond to this type of policy, therefore the post-treatment indicator will take the value of 1 after 1970. The estimation includes municipality fixed effects to control for both observed and unobserved municipality characteristics with time-invariant effects on fertility and year-fixed effects, which control for time-varying factors affecting fertility rates in the country in the same manner (for example, it accounts for the downward trend of fertility across all municipalities as shown in Fig. 3.10). In some specifications I also include municipality-specific time trends allowing for differential time-trends as suggested by Jaeger et al. (2020). Other unobserved factors are captured with the random error term $\epsilon_{m,t}$. The standard errors are robust and clustered at the municipality level to account for the possibility of serial correlation in the error term.

A potential concern from the previous specification is the definition of treatment status as using the median of the distribution is a somewhat arbitrary threshold. This is even more crucial when the treatment is not clear on and off treatment, as in the case of the radio. A recent paper by Callaway, Goodman-Bacon and Sant'Anna (2021) suggests that using a binary treatment when the treatment is continuous can bias the results due to treatment effect heterogeneity and that the bias could be am-

biguous. To deal with this issue and following Lindo, Myers, Schlosser and Cunningham (2020), I allow the treatment to vary in dosages to see if there were nonlinearities in treatment effects. More importantly, this model allows me to compare the effect of radio signal strength on fertility at different points in the radio signal distribution providing a more robust estimation. In this case, the estimation of the effect of signal strength on fertility rate corresponds to the Eq. (3.3):

$$Y_{m,t} = \alpha_m + \theta_t + \beta \text{signal}_m \times post + \epsilon_{m,t} \tag{3.3}$$

Where $Y_{m,t}$ is some measure of fertility in municipality m in year t and $signal_m$ is a set of signal strength ranges. The estimation includes time and municipality fixed effects and the standard errors are robust and clustered at the municipality level.

3.5.1 Testing knowledge and access to family planning

It is possible that the radio campaigns had an effect only on those municipalities that had a radio strong signal (knowledge) and that at the same time had *access* to a Profamilia clinic. To evaluate this, I limit the sample of municipalities to those that were not further away than 70 kilometres from the city with a Profamilia clinic. I exclude the cities that had the clinic because it is likely that urban centres differ considerably from their rural neighbours. With this sample of municipalities, I can compare places that had access to a Profamilia clinic, but that had different strengths in radio due to the orographic characteristics of the terrain.[17] I repeat the estimations of Eq. (3.2) and Eq. (3.3) for this sample of municipalities.

[17]Fig. 3.7 shows the predicted signal strength of Profamilia transmitters to the neighbouring municipalities and Table 3.10 presents the summary statistics for these municipalities.

3.6 Results

Table 3.2 shows the estimated effects of exposure to the Profamilia radio campaign on the births per women aged 15 to 48 following Eq. (3.1). Column 1 includes only the predicted radio signal strength and department fixed effects and the coefficient suggests that higher signal strength, that is an increase in exposure to the radio programme, is associated with a decrease in fertility rates. The effect is almost insignificant in economic terms: an increase of 1 standard deviation in the signal strength is associated with a reduction of 37.5 births per 1,000 women. When the hypothetical free-signal space is included, the size of the coefficient decreases as well as its significance and the effect found is insignificant and also close to zero. The results in Column 3 corroborate this finding. Once geographical, historical and population controls are added into the model, the coefficient remains small and non-significant. The results imply that there were no effects of the radio announcements promoting the family clinics on the number of births per women aged 15 to 48. The results in Appendix Fig. 3.11 confirm that the effects were non-significant and small even in municipalities with the strongest radio signal.

Table 3.2: Estimated effects of signal strength on births per 1,000 women

	births per women		
	(1)	(2)	(3)
radio signal strength	-2.58***	-0.46	-0.53
	(0.15)	(0.89)	(0.57)
Departament fe	Y	Y	Y
Free space signal		Y	Y
Geographical controls			Y
Historical controls			Y
Urbanisation			Y
Population controls			Y
Adj R^2	0.26	0.27	0.60
Num. obs.	939	939	937
N Clusters	20	20	20
Mean	612	612	612
sd	111.3	111.3	111.3

Notes: Estimates are based on a OLS model. The outcome variables are birth rates, calculated as the proportion of children between age 0 and 3 to women between ages 15-48 years old in 1973. The explanatory variable is radio signal strength. Free-space signal is the hypothetical signal strength in the free-space (i.e. assuming no topography). Geographic controls include altitude and size of the municipality. Historical controls include year of foundation of the municipality. Urbanization and population controls such as sex-ratio are measured at municipal level in 1973. Bootstrap standard errors in parentheses are clustered at the departmental level, and $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ indicate statistical significance.

A similar picture is shown by the results of the DID analysis in Table 3.3. Column 1 reports the unconditional correlation between the Total Fertility Rate and receiving a signal strength above the median signal strength. The correlation is negative, corroborating, as shown in Fig. 3.3, that on average the TFR in places with a stronger signal is lower over the period. In Columns 2 to Column 5, I present the results from the average post-treatment difference, that is, the average effect in the municipalities after the campaign started. The results in Column 2 show that the average post-treatment effect on the Total Fertility Rate is non-significant and very close to zero, which indicates a null effect of the campaign on the TFR. When we observe the Age-Specific Fertility Rates for women between 15 to 19, 30 to 34 and 40 to 44, the results in Columns 3 to 5 establish that this null effect is also present in the different age groups despite the potential differences in fertility behaviour at

different points in the fertility life of women. This again corroborates that the radio campaign did not affect fertility rates.

Table 3.3: Estimated effects of radio signal strength on fertility - binary treatment

	TFR (1)	TFR (2)	ASFR 15-19 (3)	ASFR 30-34 (4)	ASFR 40-44 (5)
Strong signal	-0.56***				
	(0.04)				
Strong signalxPost		0.052	0.003*	-0.002	0.000
		(0.04)	(0.002)	(0.003)	(0.003)
(Intercept)	6.98***				
	(0.03)				
Mean	6.7	6.7	0.12	0.28	0.13
sd	1.5	1.5	0.05	0.08	0.08
Municipality fe	N	Y	Y	Y	Y
Year fe	N	Y	Y	Y	Y
Adj. R^2	0.04	0.87	0.71	0.66	0.65
Num. obs.	5,514	5,514	5,488	5,501	5,470
N Clusters	513	513	513	513	513

Notes: Estimates are based on a OLS model evaluating expected fertility rates in Colombian municipalities excluding cities were the transmitters were located. Strong signal_m equals 1 if the signal strength of the municipality is above the median signal strength. Standard errors in parentheses are clustered at the municipality level, and $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ indicate statistical significance.

However, a potential concern from the previous results is the definition of the treatment status as a dichotomous variable given that signal strength is a continuous variable. To overcome this, I allow the treatment to vary in dosages grouping the signal strength into 5 groups. In this specification, the average post-treatment effect is the differential fertility after 1970 between the group of municipalities with the weakest signal strength (-112.5 dBm to -44.2 dBm) and municipalities under different ranges of signal strength. In general, the results from Table 3.4 show non-significant and economically small effects very close to zero which confirms what has been previously discussed. Interestingly, the results in Column 1 suggest that in comparison to the municipalities with the weakest signal strength the municipalities with the strongest signal reception (-20.1 dBm to 6.1 dBm) experienced a slower fertility decline. The difference, however, is minimal. The results imply that although

fertility was changing rapidly after 1970 across the territory, these changes were not associated with radio announcements.

Table 3.4: Estimated effects of radio signal strength on fertility - continuous treatment

	TFR (1)	ASFR 15-19 3 (2)	ASFR 30-34 (3)	ASFR 40-44 (4)
$[-112.5, -44.2)dBm$xPost (ref. group)				
$[-44.2, -36.5)dBm$xPost	0.12	0.01^{**}	0.01^{*}	-0.00
	(0.07)	(0.00)	(0.00)	(0.01)
$[-36.5, -29.1)dBm$xPost	0.10	0.01^{**}	0.01	0.00
	(0.07)	(0.00)	(0.00)	(0.00)
$[-29.1, -20.1)dBm$xPost	0.10	0.00	-0.00	-0.00
	(0.07)	(0.00)	(0.00)	(0.00)
$[-20.1, 6.1)dBm$xPost	0.19^{**}	0.01^{***}	0.01^{*}	-0.00
	(0.07)	(0.00)	(0.00)	(0.01)
Mean	6.7	0.12	0.28	0.13
sd	1.5	0.05	0.08	0.08
Municipality fe	Y	Y	Y	Y
Year fe	Y	Y	Y	Y
Adj. R^2	0.78	0.64	0.56	0.57
Num. obs.	5,628	5,602	5,615	5,584
N Clusters	936	936	936	936

Notes: Estimates are based on a OLS model evaluating expected fertility rates in Colombian municipalities excluding cities were the transmitters were located. The excluded category is signal strength between -112.5 dBm and -44.2 dBm. Standard errors in parentheses are clustered at the municipality level, and $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ indicate statistical significance.

3.6.1 Municipalities with access to a clinic

The previous results raise the potential concern that when looking at women in the whole country, the limited effects of the radio campaign could be explained by limited access to a Profamilia clinic, as the clinics were located far away from several municipalities. This is also a mechanism through which family planning could have affected fertility: not only through knowledge of availability but also through access. In Table 3.5 and Table 3.6 I report the result for the estimated effects of radio signal strength on fertility for women living in municipalities located less than 70 km away from cities that had a Profamilia clinic by 1970 following Eq. (3.2) and Eq. (3.3).[18] All

[18]I considered 70 km because this distance allows people to move from the centre of a municipality to another in approximately 1 hour and a half in public transportation. Table 3.11 and Table 3.12 show the results changing the distance from the centre of the city for 30 km and 140 km. In both cases, there is no effect of the radio programme on fertility.

of these municipalities had access to a clinic while experiencing differences in signal strength due to topographic factors.

The results from Table 3.5 show that even when limiting the sample to municipalities near a clinic, places that received a stronger radio signal did not experience a faster fertility decline one year after the radio campaign started. The results in Columns 2 to Column 4 that report the average post-treatment difference indicate that the effect on the different fertility rates was non-significant and economically marginal. This implies that radio exposure did not have a heterogeneous effect on fertility behaviour in women of different ages. Women between 15 to 19 years old that were exposed to the campaign and lived in a municipality close to a Profamilia clinic, did not reduce their fertility in comparison to women of the same age group that received a weaker radio signal.

In Table 3.6 I address again the issue regarding the threshold of the treatment status accounting for differences in the intensity or dosage of the radio signal. Once again, the results confirm that there is no effect of the intensity of the radio signal after the start of the campaign on fertility rates, either by looking at the TFR or by looking at different age-groups fertility. Overall, municipalities that received a stronger radio signal (-15.3 dBm to 6.1 dBm) did not experience lower fertility one year after the start of the radio campaign compared to those municipalities that received the weakest signal strength (-47.8 dBm to -32.4 dBm), even though these group of municipalities had access to a Profamilia clinic.

Table 3.5: Estimated effects of radio signal strength on fertility - binary treatment

	TFR (1)	ASFR 15-19 (2)	ASFR 30-34 (3)	ASFR 40-44 (4)
Strong signalxPost	0.03	0.000	0.002	-0.004
	(0.04)	(0.002)	(0.003)	(0.003)
Mean	6.5	0.11	0.27	0.13
sd	1.3	0.05	0.07	0.07
Municipality fe	Y	Y	Y	Y
Year fe	Y	Y	Y	Y
Adj. R^2	0.86	0.70	0.70	0.64
Num. obs.	3,078	3,072	3,078	3,070
N Clusters	513	513	513	513

Notes: Estimates are based on an OLS model evaluating expected fertility rates in municipalities less than 70 km away from cities that had a Profamilia clinic by 1970. Strong signal_m equals 1 if the signal strength of the municipality is above the median signal strength. Standard errors in parentheses are clustered at the municipality level, and $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ indicate statistical significance.

Table 3.6: Estimated effects of radio signal strength on fertility - continuous treatment

	TFR (1)	ASFR 15-19 3 (2)	ASFR 30-34 (3)	ASFR 40-44 (4)
$[-47.8, -32.4)dBm$xPost (ref. group)				
$[-32.4, -26.4)dBm$xPost	−0.08	−0.04	−0.02	−0.00
	(0.08)	(0.02)	(0.02)	(0.01)
$[-26.4, -20.8)dBm$xPost	−0.13	−0.04	−0.02	0.01
	(0.07)	(0.02)	(0.01)	(0.01)
$[-20.8, -15.3)dBm$xPost	0.01	−0.04	−0.03	0.01
	(0.07)	(0.02)	(0.01)	(0.00)
$[-15.3, 6.1)dBm$xPost	−0.01	−0.04	−0.02	0.01
	(0.07)	(0.01)	(0.01)	(0.01)
Mean	6.5	0.11	0.27	0.13
sd	1.3	0.05	0.07	0.07
Municipality fe	Y	Y	Y	Y
Year fe	Y	Y	Y	Y
Adj. R^2	0.76	0.62	0.59	0.56
Num. obs.	3,078	3,078	3,078	3,078
N Clusters	513	513	513	513

Notes: Estimates are based on an OLS model evaluating expected fertility rates in municipalities located less than 70 km away from cities that had a Profamilia clinic by 1970. The excluded category is signal strength between −47.8 dBm and −32.4 dBm. Standard errors in parentheses are clustered at the municipality level, and $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ indicate statistical significance.

3.6.2 Caveats

All the specifications pointed to a limited effect of Profamilia's radio campaign. It is possible that the information that the radio announcements were disseminating had little impact on fertility rates because the ads did not distribute particular new information or social norms, and did little to shape new behaviours or to persuade couples to limit their fertility. Another reason could be that the radio announcements are broadcast when the fertility decline is already on its way across the whole country and therefore the information about contraception and fertility control is not new.

However, three caveats should be taken into consideration to interpret these results. First, the timing of the campaign and the limited effects that I found in the previous section could indicate that the campaign did not spread any new information that could result in changes in behaviour or increase the adoption of contraceptive methods. Unfortunately, the effects on general contraceptive usage are still largely unknown since there is no systematic information on utilisation in the census data. However, the statistics collected by Simmons and Cardona (1974) and shown in Table 3.1 confirm that by 1969 more than 60% of the women in the country, including women living in rural areas were favourable to family planning. Furthermore, the positive reaction that the public had to the radio campaign as suggested by the director of Profamilia indicates that fertility control was already known and accepted in the country, even before the start of the radio campaign.

Second, I am evaluating the short-term effects on the fertility of the radio campaign. Because the census was carried out 4 years after the release of the campaign I can only provide fertility estimations for the period 1966 to 1971 and post-treatment effects for 1971. The reason why I do not include the TFR of 1972 is that this estimation is likely to be biased downward due to under-registration or misreporting of ages, as discussed in the previous chapter.

Additionally, estimating the effects of the radio campaign in the long term could be challenging because of the other potential interventions that could affect fertility. For instance, in 1973 Profamilia started another radio campaign in alliance with one of the most important radio stations in the country: Radio Sutatenza. The station, which was the first Catholic radio station in Latin America, was originally founded in the town of Sutatenza in the department of Boyaca in 1947 with the idea of combating adult illiteracy. In 1973, with the most powerful transmitters and with national coverage the radio station launched the campaign "Responsible Procreation" with several radio spots that lasted between twenty to sixty seconds several times a day between 5:00 am and midnight (Roldán, 2014). As the previous Profamilia campaign, the campaign did not explicitly recommend a particular form of birth control but instead promoted mutual respect and shared decision-making between spouses as core values of the Christian marriage. The spots were supported with newspapers, radio dramatisations of 15 minutes long and theatre scripts. The campaign defended the idea of having a child as a deliberate choice in which both spouses had equal say but it also extend health education (Roldán, 2014).

Yearly effects estimations

A third concern emerges regarding the moment in which the radio campaign should affect fertility. Although the campaign started in mid-1969, it is expected that the reaction of fertility would not be immediate and possibly only measurable after around 9 months or so. Therefore, to observe the effects of the radio campaign by each year, from 1966 to 1971, and to test the potential lag in response to the radio campaign I estimate the following model in equation Eq. (3.4):

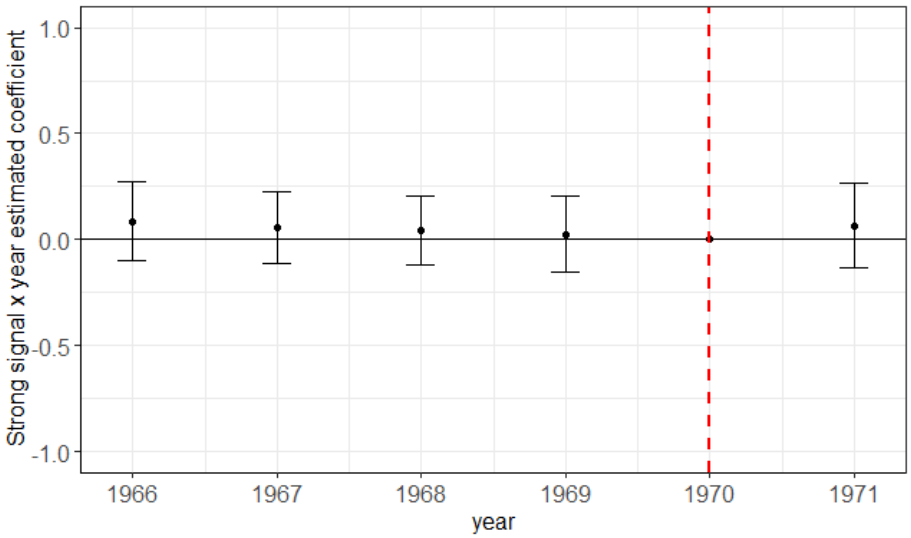
$$Y_{m,t} = \beta_1 \text{strong signal}_m + \beta_2 \text{DT}_t + \beta_3 \text{strong signal}_m \times \text{DT}_t + \alpha \text{DM}_t + \epsilon_{m,t} \quad (3.4)$$

Where $Y_{m,t}$ is some measure of fertility, strong signal$_m$ equals 1 if the signal strength of the municipality is above -32.4 dBm, which corresponds to the lower quintile of the radio strength distribution, DT$_t$ is a year dummy and DM$_m$ is a municipality dummy.[19] As in the other estimations, standard errors were clustered at the municipal level allowing for arbitrary serial correlation and our main coefficient of interest is β_3.

Fig. 3.4 presents the plotted coefficients from Eq. (3.4) where the outcome of interest is the Total Fertility Rate in each year. As can be seen, there is no significant effect of the radio campaign on the total fertility rate after the radio campaign started. In other words, places with a stronger signal strength did not experience lower fertility rates one year after the campaign was introduced. The results also suggest that there were no significant differences between places with weak and strong signal strength before the broadcasting of the radio campaign i.e. there are no pre-trends in the data. Fig. 3.5 shows the results for age-specific fertility rates for the different age groups and confirms that the radio campaign did not have an effect on fertility after the start of the radio announcements and that there were no significant pre-trends.

[19] I set the threshold at the lower quintile instead of that in the median because this allows me to compare municipalities that got the worse signal to those that received some or good signal.

Figure 3.4: Estimated effects of radio signal strength on Total Fertility Rates

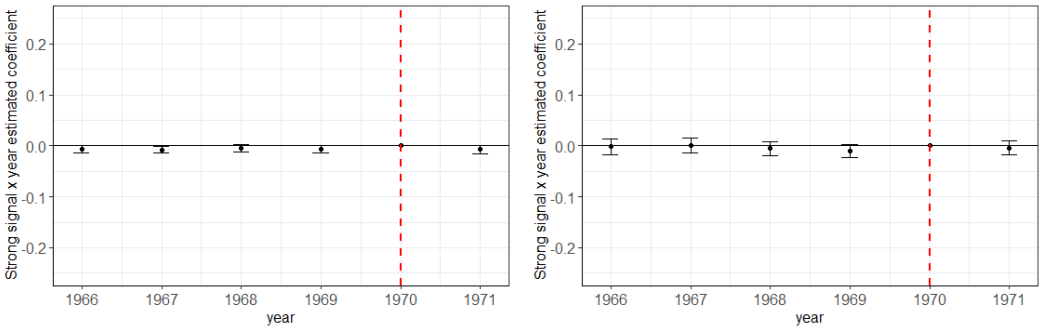


Note: The figure shows the coefficients and standard errors from DD regression Eq. (3.4). Estimates are based on an OLS model evaluating expected fertility rates in municipalities living less than 70 km away from cities that had a Profamilia clinic by 1970. Strong signal$_m$ equals 1 if the signal strength of the municipality is above -32.4 dBm, which corresponds to the lower quintile of the radio strength distribution. Standard errors clustered at the municipal level.

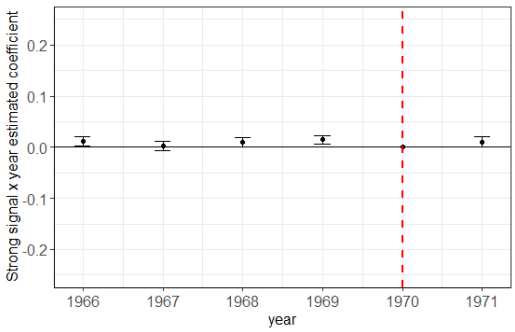
Figure 3.5: Effects of radio signal strength on Age-Specific Fertility Rates

(a) 15 - 19 years old

(b) 30 - 34 years old



(c) 40 - 44 years old



Note: The figures show the coefficients and standard errors from DD regression Eq. (3.4). Fig. 3.5a shows the estimated effects of radio on the age-specific fertility rate of women between 15 to 19 years old. Fig. 3.5b shows the effects on the age-specific fertility rate of women between 30 to 34 years old. Fig. 3.5c shows the effects on the age-specific fertility rates of women between 40 to 44 years old to measure the effect of the campaign on stopping. Estimates are based on an OLS model evaluating expected fertility rates in municipalities living less than 70 km away from cities that had a Profamilia clinic by 1970. Strong signal$_m$ equals 1 if the signal strength of the municipality is above -32.4 dBm, which corresponds to the lower quintile of the radio strength distribution. Standard errors clustered at the municipal level.

3.7 Conclusions

This chapter investigates if a national radio campaign by Profamilia promoting family planning clinics in the late 1960s in Colombia accelerated the fertility decline across the country. Some argue that contraception was, without question, the main force underlying the fertility decline of developing economies (Weinberger, 1987). But for Colombia, Miller shows that, in urban areas, the effects of Profamilia on lifetime

fertility were modest. I focus on the effects of the Profamilia radio campaign on fertility, as mass media increases the availability of information, especially in rural areas. Additionally, mass media has the potential to shape new social roles, change social norms and raise the level of education of the population.

To measure the effectiveness of the radio campaign on fertility this chapter exploited exogenous variation in radio signal strength that results from topographic factors. Using professional radio software I predicted the radio signal strength of a traditional radio station that was located in the cities where Profamilia launched its radio campaign. With the predicted signal strength, I estimate a lower bound of the intention to treat the effect. As a first strategy, I regressed fertility rates on the predicted radio signal strength while controlling for the hypothetical signal strength where there are no orographic obstacles such as a mountain. Additionally, I employed a difference-in-differences strategy to compare fertility rates before and after the start of the radio campaign. Finally, I test two mechanisms through which the campaign could have affected fertility: through knowledge of availability or through access to clinics and contraceptive methods. To do so, I restricted the sample of municipalities to those not further away than 70 kilometres from a city with a family clinic.

The results from this chapter show that the Profamilia radio campaign had a limited effect on fertility rates. Although several evaluations of the period suggested that the announcements were successful in increasing the number of visitors to the clinics, the findings in this chapter demonstrate that the overall short-term effect of the campaign was limited, as it was not translated into a decrease in fertility rates. Overall, places with a stronger signal strength did not experience lower fertility rates one year after the campaign was introduced. The test for the mechanisms reveals that not even places that were close to a Profamilia clinic experienced faster fertility. Miller (2010) had previously shown that access to family planning had a minor effect

on the rapid fertility decline of the country. My results confirm this finding and show that knowledge of family planning also had a moderate effect.

As the radio campaign did little in disseminating new social roles or raising the economic aspirations of the population, the results suggest that the potential effects of media on fertility rely more on the promotion of new social norms than on the diffusion of information about family planning. Similar to what the literature has found, these findings suggest that most of the effects of radio on socioeconomic outcomes are explained by the capacity of persuasion and not only by the diffusion of information (Dellavigna & Gentzkow, 2010; La Ferrara et al., 2012).

The modest role of Profamilia in the Colombian fertility decline suggests that other forces explain the fertility decline. In particular, changes in the demand for children could be the driver of the fertility decline. Jayne and Guilkey (1998) and Miller (2010) suggest that the single most important factor in the fertility decline in Colombia was not the availability of contraceptives but access to education. In the next chapter, I focus on the long-run relationship between fertility and women's education.

3.A Appendices

3.A.1 Family planning and radio

This appendix provides additional information regarding exposure to the radio.

By the end of the 1960s radio was very popular across the country with close to 20 million listeners in urban as well as rural areas. However, the information on how many people had a radio at home or listened to the radio is not available. Table 3.7 presents statistics compile by Stycos and Avery (1975) for 1971 on the exposure to the radio of new visitors of different Profamilia clinics. The table shows that on average

89% of women that went to a Profamilia clinic during July and August of 1971 listen to the radio, 40% of these women listen to the radio more than 36 hours per week, and 83% heard one of Porfamilia's spots.

Table 3.7: Exposure of new clients to radio and Profamilia announcements, 1971

City	**Listen to Radio (%)**	**Listen 36 Hrs. + per week (%)**	**Heard Profamilia announcement (%)**
Bogotá	97	30	96
Medellín	97	55	89
Cali	64	22	44
Barranquilla	87	71	86
Bucaramanga	84	32	74
Manizales	98	44	95
Pereira	90	47	86
Cúcuta	79	27	74
Ibagué	94	78	94
Palmira	94	45	91
Pasto	90	20	82
Armenia	89	26	80
Neiva	90	44	91
Total	**89**	**40**	**83**

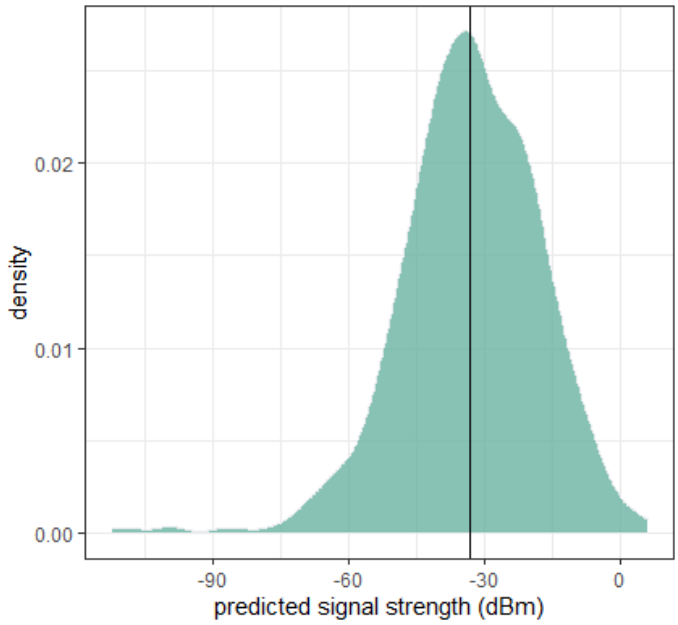
Notes: The data comes from Stycos and Avery and shows the results of interviews of women who attended the clinics during July and August 1971.

3.A.2 Data

This appendix presents additional information on the data collected and estimated. First, it shows the distribution of the predicted signal strength for the whole country. Then I show the predicted signal strength calculated for each neighbouring municipality that is less than 70 km away from the transmitter of *Radio Reloj*. I also present the distribution of the Crude Birth Rate for the whole country, and the Total Fertility Rate by quintiles for the whole country and for each neighbouring municipality.

Signal strength

Figure 3.6: Distribution of the predicted signal strength



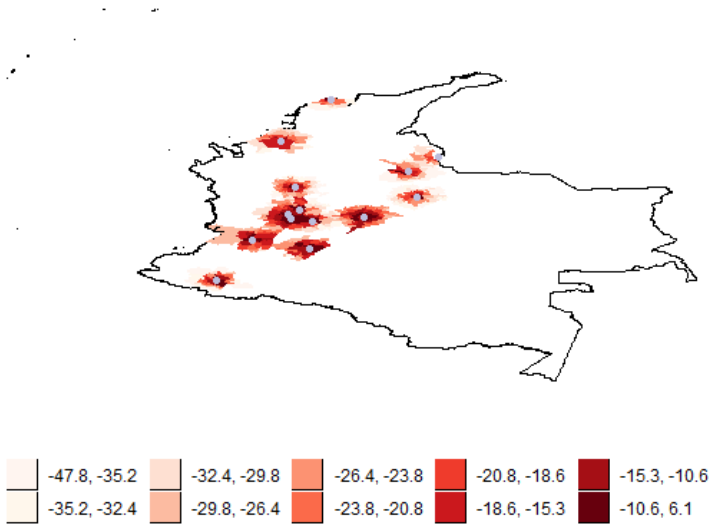
Notes: The graph shows the distribution of the predicted signal strength of *Radio Reloj*, computed using the Irregular Terrain Model (ITM) and measured in decibel-milliwatts (dBm) using the complete sample. The black line is the median of the distribution. Sources: Authors' calculations based on Miller; Bailey; Stycos and Avery and the National directory of broadcasting (1976).

Table 3.8: Descriptive statistics predicted radio signal

	Min	Median	Mean	Max	sd
Effective signal strength	-112.48	-32.79	-33.15	6.09	15.66

Notes: Descriptive statistics of the predicted signal strength of *Radio Reloj*, computed using the Irregular Terrain Model (ITM) and measured in decibel-milliwatts (dBm) using the complete sample. Sources: Authors' calculations based on Miller; Bailey; Stycos and Avery and the National directory of broadcasting (1976).

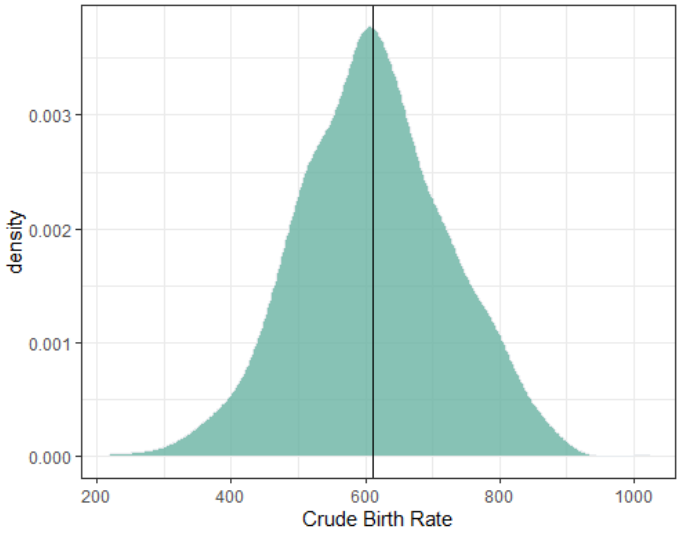
Figure 3.7: Location and signal strength of Profamilia transmitters - municipalities close to a Profamilia clinic



Notes: The map shows the maximum predicted signal strength in each neighbouring municipality that is less than 70 km away from the transmitter of *Radio Reloj* and it is computed using the Irregular Terrain Model (ITM) and measured in decibel-milliwatts (dBm). The grey dots are the transmitter location. Sources: Authors' calculations based on Miller; Bailey; Stycos and Avery and the National directory of broadcasting (1976).

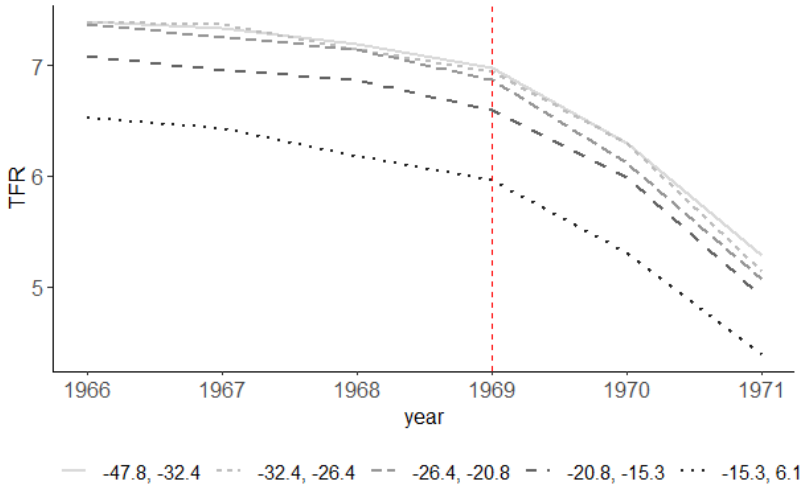
Fertility

Figure 3.8: Distribution of the General Fertility Rate



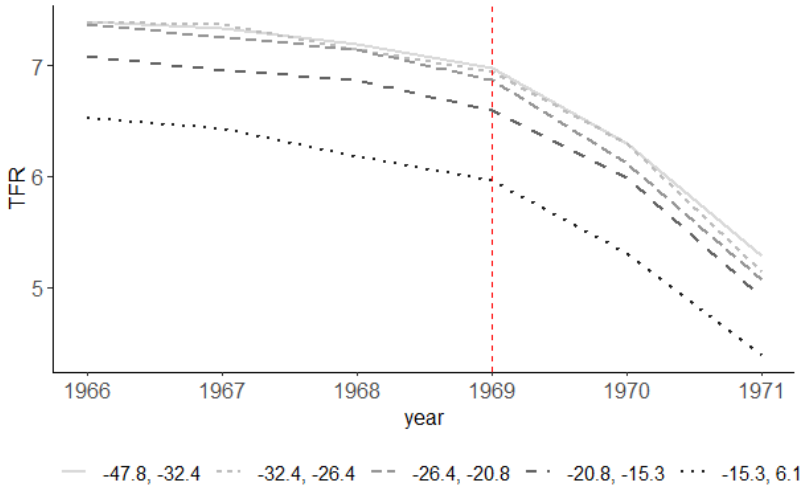
Notes: The graph shows the distribution of the General Fertility Rate, computed using children between 0 to 3 years old and women between 15 to 48 years old. The black line is the median of the distribution. Sources: Authors' calculations based on Miller; Bailey; Stycos and Avery and the National directory of broadcasting (1976).

Figure 3.9: TFR 1966-1971, National



Note: The figure shows the Total Fertility Rate (TFR) for municipalities in Colombia grouping them based on the strength of the signal radio they receive. Calculations using the OCM. Women between 15 and 64 years old and children between 0 and 7 years old. Sources: Authors' calculations based on 1973 full Census data.

Figure 3.10: TFR 1966-1971, Municipalities close to a Profamilia clinic



Note: The figure shows the Total Fertility Rate (TFR) for municipalities that are less than 70 km. away from the transmitter. Calculations using the OCM. Women between 15 and 64 years old and children between 0 and 7 years old. Sources: Authors' calculations based on 1973 full Census data.

Table 3.9 presents summary statistics dividing the municipalities in the country into those that received signal strength above the median with those that received a signal strength below the median. On average municipalities with a strong signal had lower

fertility, were more populated and were closer to a Profamilia clinic.

Table 3.9: Summary statistics by signal strength - All country

	Strong signal	**Weak signal**
Total Fertility Rate in 1968	6.81	7.48
	(1.24)	(1.44)
Crude Birth Rate in 1970-1973	584	641
Population in 1973	28,173	14,004
Urbanization rate	36%	32%
Distance to Profamilia city (kms.)	42.15	113.70
Sex ratio	0.88	0.98
Number of municipalities	468	470

Notes: Summary statistics for municipalities of Colombia. Strong signal strength means that the municipality received a signal strength that is above the median. The Total Fertility Rate is computed using the Own Child Method. The Crude Birth Rate is the proportion of children age 0 to 3 years old in the census to women ages 15 to 48. Sex ratio is the proportion of men ages 15 to 50 to women ages 15 to 48. Source: Authors' calculations based on 1973 full Census data.

Table 3.10 presents summary statistics dividing the neighbouring municipalities of cities with a Profamilia clinic by 1970 into those that received signal strength above the median with those that received signal strength below the median. On average municipalities with a strong signal had lower fertility, were more populated and were closer to a Profamilia clinic.

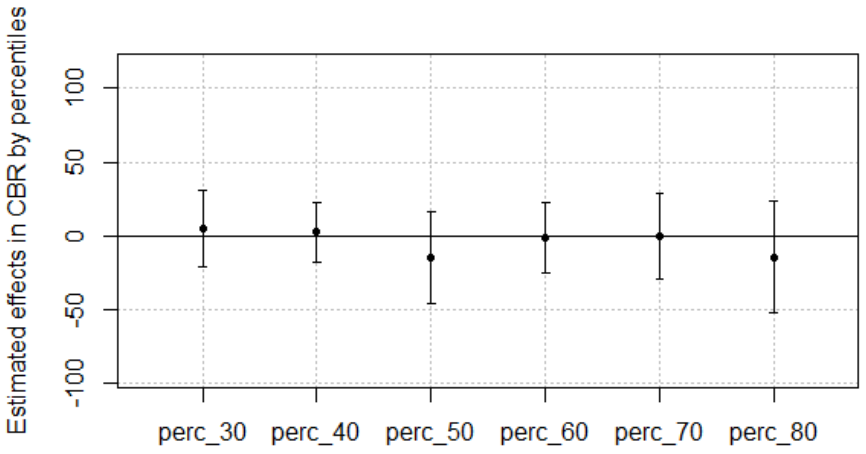
Table 3.10: Summary statistics by signal strength - municipalities with access to clinic

	Strong signal	**Weak signal**
Total Fertility Rate in 1968	6.64	7.17
	(1.08)	(1.23)
Crude Birth Rate in 1970-1973	568	618
Population in 1973	15,539	11,555
Urbanization rate	38%	31%
Distance to Profamilia city (kms.)	31.12	54.11
Sex ratio	1.04	1.03
Number of municipalities	248	265

Notes: Summary statistics for neighbouring municipalities of cities with a Profamilia clinic by 1970. Strong signal strength means that the municipality received a signal strength that is above the median. The Total Fertility Rate is computed using the Own Child Method. The Crude Birth Rate is the proportion of children age 0 to 3 years old in the census to women ages 15 to 48. Sex ratio is the proportion of men ages 15 to 50 to women ages 15 to 48. Source: Authors' calculations based on 1973 full Census data.

3.A.3 Robustness

Figure 3.11: Effects on births per women aged 15 to 48 by percentiles of signal strength



Note: The figure shows the coefficients and standard errors from Eq. (3.1) where $signal_m$ is measured using dummy variables corresponding to different percentiles of the signal. Estimates are based on an OLS model. The outcome variables are birth rates, calculated as the proportion of children between age 0 and 3 to women between ages 15-48 years old in 1973. Geographic controls include altitude and size of the municipality. Historical controls include year of the foundation of the municipality. Urbanization and population controls such as sex-ratio are measured at the municipal level in 1973. Standard errors in parentheses are clustered at the departmental level.

Table 3.11: Effects of radio signal strength on fertility (30 km away from clinic)

	TFR	TFR	ASFR 15-19	ASFR 30-34
Strong signal	-0.81***			
	(0.09)			
Strong signalxPost-1970		-0.03	0.0016	-0.0013
		(0.08)	(0.0033)	(0.0051)
(Intercept)	6.493***			
	(0.06)			
Municipality fe	N	Y	Y	Y
Year fe	N	Y	Y	Y
Adj. R^2	0.098	0.88	0.21	0.32
Num. obs.	780	780	3,572	3,861
N Clusters	130	130	130	130

Notes: Estimates are based on an OLS model evaluating expected fertility rates in municipalities living less than 30 km away from cities that had a Profamilia clinic by 1970. Strong signal_m equals 1 if the signal strength of the municipality is above the median signal strength. Standard errors in parentheses are clustered at the municipality level, and $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ indicate statistical significance.

Table 3.12: Effects of radio signal strength on fertility (140 km away from clinic)

	TFR	TFR	ASFR 15-19	ASFR 30-34
Strong signal	-0.50***			
	(0.04)			
Strong signalxPost		0.04	0.0018	-0.0031
		(0.0385)	(0.0016)	(0.0026)
(Intercept)	6.906***			
	(0.03)			
Municipality fe	N	Y	Y	Y
Year fe	N	Y	Y	Y
R^2	0.031	0.85	0.21	0.27
Observations	4,974	4,974	22,636	24,501
N Clusters	829	829	829	829

Notes: Estimates are based on an OLS model evaluating expected fertility rates in municipalities living less than 140 km away from cities that had a Profamilia clinic by 1970. Strong signal_m equals 1 if the signal strength of the municipality is above the median signal strength. Standard errors in parentheses are clustered at the municipality level, and $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ indicate statistical significance.

4

Can female education explain the fertility decline?

4.1 Introduction

Across the world educated women tend to have fewer children than their less-educated peers. There are several mechanisms behind this association and distinguishing between these channels has proven challenging. Caldwell (1980) argues that "the primary determinant of the timing of the onset of the fertility transition is the effect of mass education on the family economy" (p.225). Caldwell (1980) suggests that once mass education is introduced in any given society (and it is profitable) families face the new cost of educating their children and therefore are forced to reduce their fertility. It is then expected that the first generation that enjoyed mass schooling will lead to a fertility decline. Similar to this, G. Becker and Lewis (1973) argue that parents maximise a utility function of quantity and expenditure on children (quality). The model implies that an increase in income and higher rates of returns in education explains the decision of parents to prefer fewer and better-educated children (see also Galor and Weil (2000)).[1]

Apart from increasing the cost of children, education can also increase the opportunity costs of having children for women (Liu & Raftery, 2020; Frye & Lopus, 2018). At the same time, mass education can speed up cultural change, create new cultures and values, and raise the economic aspirations of the population (Caldwell, 1980). For example, more educated women have a higher demand for and greater use of family planning. Nonetheless, the empirical evidence testing these mechanisms is limited given that other variables, such as wealth or family preferences, are strongly related both to schooling and fertility, which creates confounding bias. Additionally, the simultaneity in the decision-making between childbearing and schooling limits the causal interpretation of the relationship (Behrman, 2015).

[1]The empirical literature on the existence of a quality-quantity tradeoff finds mixed results, especially when it studies the historical fertility transition (Guinnane, 2011; S. Becker et al., 2010; Bleakley & Lange, 2009; Tan, 2019).

Furthermore, the effects of women's education on fertility can change depending on how fertility is defined. Most of the studies focus on the number of children a woman has (intensive margin) while relatively few studies include the decision of ever having a child (extensive margin) (Aaronson, Lange & Mazumder, 2014; Baudin, De La Croix & Gobbi, 2015; De La Croix, Schneider & Weisdorf, 2019). Additionally, differences in fertility measures do not allow us to draw inter-temporal comparisons and understand how the relationship between female education and fertility changes over time.

This chapter provides new stylised facts about the long-run relationship between women's education and fertility at several levels of aggregation and along the two margins of fertility.

The case of Colombia is of relevance as the country adopted a set of liberal reforms that placed improvements in education as one of the main priorities after the 1930s. At the same time, these reforms aimed at promoting gender equality by making access to education equal for girls and boys. These reforms represented a break with the Catholic traditional views, which dominated the educational system since the late 19^{th} century and limited the participation of women in the labour market.

I use individual-level data from the complete census of 1973 and individual-level data from a 10% sample of the censuses of 1985 and 1993 from the Integrated Public Use Microdata Series – IPUMS International (Minnesota Population Center, 2019) to estimate the national Total Fertility Rate by educational level from 1958-1990. I document that fertility declined across all educational groups. Also, the fertility gap between the lowest and highest educated women existed during the onset of the fertility transition and the gap remained virtually unchanged by 1990. Fertility declined faster for women with secondary education, from 6 children to 2 between 1960 to 1990, whereas for women with no education fertility dropped from 8 children to 4

between the same years.

Then I turn my attention to the intensive and extensive margin by studying the fertility decisions of women with completed fertility. The results at the national level suggest that in the intensive margin, the fertility of women (conditional on being mothers) decreases with education. For women born after 1930 the number of children born decreased across all educational groups, but the decline was less pronounced for women with no education. In the extensive margin, I find a fall in childlessness in married women across all educational groups, which implies that fewer married women stayed childless. On the contrary, the levels and trend of childlessness in single women were more or less constant during the fertility transition. Around 90% of single women with secondary education remained childless but 50% of single women with no education were mothers. This characterisation of the fertility transition differs considerably from the European case, as it reveals high levels of illegitimacy for uneducated women.

Next, I focus on the relationship between fertility and female education at the sub-national level. I show that across the country, education shares grew while fertility declined. However, from 1960 to 1970, changes in fertility rates were not strongly associated with changes in female schooling rates. In other words, departments that expanded education more rapidly did not always experience greater fertility decline. Fertility decline happened broadly everywhere and the decline happened across all educational groups.

Finally, I study the relationship between schooling and fertility at the individual level, analysing both margins of fertility and the use of strategies to control fertility (starting, spacing and stopping). Given that education can potentially create spillovers from educated to uneducated women, I take into account individual as well as peer effects on fertility. The results show that having primary schooling

increases the probability of remaining childless, reduces the total number of children and reduces the probability of having a birth at a younger age for women born after 1920. For women born between 1910-1920, the effect of primary schooling on fertility is moderate. The effect of secondary schooling is stronger than the effect of primary and in all cohorts secondary schooling reduces fertility and increases the probability of remaining childless up to 124%.

Community variables, such as the percentage of peers with secondary education, have a limited effect on the total number of children, the probability of remaining childless and the use of strategies to limit fertility. This means that the externalities of education had a moderate effect on uneducated women. Therefore, individual access to education seems to be a key to fertility decisions. The results are robust to the inclusion of municipality-fixed effects.

The findings of the chapter reveal that the relationship between fertility and women's education is not always monotonic and, the relationship changes significantly depending on the level of aggregation studied. At the national and sub-national levels, the fertility decline in Colombia cannot be explained through changes in education. However, at the individual level, the results indicate that access to secondary education was key for fertility decisions, even before the fertility transition.

The results of this chapter contribute to several strands of the literature. First, they show that as female education expanded in the country and fertility declined, the gap in fertility levels between educated and non-educated women remained unchanged. Although Potter et al. (1976) showed this for 1973 and Batyra (2020) for 2005, I provide a long-run view of this relationship using consistent measures of fertility that allow for inter-temporal comparison. Additionally, I provide novel evidence on the relationship between females' education and the decision of remaining childless during the fertility transition that had not yet been studied for the Colombian case.

The Colombian case allows us to study the fertility decline while observing the expansion of education. The results reveal that fertility decline happened independently from educational expansion at the national and sub-national levels. At the individual level, the absence of peer effects also allows us to rule out the diffusion hypothesis from educated to non-educated women at the municipal level. Similarly, the results suggest a moderate importance of other economic transformations, as most of this happens at the municipal level.

4.2 Potential mechanisms

The negative relationship between fertility and education is well-documented in the literature (Castro Martin, 1995; Caldwell, 1980; Weinberger, 1987).[2] At the cross-country level, more educated countries have lower fertility rates, and at the national level, more educated women tend to have fewer children. A broad set of mechanisms can explain this relationship, but there is some agreement that women's education appears as one of the most crucial factors in fertility decline. Overall individual-level education can affect the demand and supply of children as it has the potential to change attitudes towards childbearing (Caldwell, 1980), reduce the economic utility of children (Galor & Weil, 2000; G. Becker & Lewis, 1973), increase the opportunity cost of women (G. Becker, 1960) and change their relative autonomy within the family (Jejeebhoy, 1995). However, female education could also affect fertility decisions indirectly. As Frye and Lopus (2018) highlight, the social meaning of a woman's educational attainment depends on the educational attainments of her peers which could be of particular importance during a time of women's educational expansion. In this section, I review some channels through which female education can affect fertility decisions.

[2]Recent contributions from Doepke, Hannusch, Kindermann and Tertilt (2022) argue that this regularity held in the past while it does not longer hold universally today. Mainly because in high-income countries this relationship has flattened or even reversed.

4.2.1 The intensive and extensive margins of fertility

Although fertility can be studied in two different margins, most research focuses on the intensive margin, that is the number of children a woman has. But recently some attention has been also placed on the extensive margin, that is the decision to have children or not (see for example Aaronson et al. (2014); Baudin et al. (2015); De La Croix et al. (2019)). Studying fertility at both margins is relevant as the potential mechanisms behind these decisions are different. Also, as shown by De La Croix et al. (2019), the decision to remain childless was particularly important for the reproductive strategies of pre-industrial societies. For the case of Latin America, the extensive margin is crucial as motherhood has been traditionally almost universal and early (Rosero-Bixby, Castro-Martín & Martín-García, 2009; Poston, Kramer, Trent & Yu, 1983). What is more, motherhood is largely independent of marital status, and illegitimacy rates (children outside wedlock) have been increasing since the 1960s (Esteve et al., 2016).[3]

Both margins are influenced, among other things, by the price of raising a child as well as the education of the mother (Aaronson et al., 2014; Baudin et al., 2015; De La Croix et al., 2019; Baudin, De la Croix & Gobbi, 2020). Additionally, the intensive and extensive margins can either be negatively or positively correlated. For example, Aaronson et al. (2014) shows that the expansion of schooling opportunities under the Rosenwald Rural Schools Initiative decreased the price of child quality and this was associated with a reduction in the number of children born but it also led to a decrease in childlessness rates as more women entered on the extensive margin.

As for the price of children, Aaronson et al. (2014) predict that fewer women will remain childless as longevity increases, and more women will choose to have children in response to an improvement in health conditions. Similar reasoning implies

[3]Therefore, the Latin American case differs considerably from the European case that had low levels of illegitimacy during the fertility transition (Clark, 2007).

that fewer women should remain childless as the costs of education decline.

The relationship between women's education and childlessness potentially occurs because the opportunity cost of bearing and rearing a child is higher for higher-educated women, for example, more-educated women have presumably more labour opportunities than less-educated women, facing a higher trade-off. Looking at the US for over 100 years, Baudin et al. (2015) show that voluntary childlessness was the main driver of childlessness until education increased the opportunity cost for parents.

In the pre-industrial context, De La Croix et al. (2019) show that the English upper classes were more often childless than the rest, which resulted in comparatively modest rates of reproduction among the upper classes. The authors conclude that different factors explain these variations. In a time before effective contraception within marriage, it was difficult for couples who were sexually active to avoid pregnancy so they suggest that sexual abstinence could vary across classes. Also, the authors argue that it is possible that higher rates of venereal diseases such as gonorrhoea, chlamydia and syphilis among the upper classes led to higher rates of sterility.

4.2.2 Individual-level effects: women's educational attainment

A causal examination of the relationship between fertility and female's education is often limited by the simultaneity in the decision-making between childbearing and schooling. Several papers have addressed this issue by looking at the implementation of universal education policies or the introduction of compulsory schooling (Behrman, 2015; Osili & Long, 2008; Alzúa & Velázquez, 2017). Overall, these studies find that increases in education significantly reduce fertility and women's desired family size, and some have suggested changes in quality or quantity as the main mechanism

(Bleakley & Lange, 2009).

Identifying changes in the quality of children is often limited by the fact that compulsory education does not always translate into universal or free education, as in the case of Colombia. Also, the evidence suggests that the expansion of free schooling opportunities decreases the price of child quality which can lead to a decrease in childlessness rates (Baudin et al., 2015). Others find that an increase in education reduces overall fertility. Cortés, Gallego and Maldonado (2016) show that educational conditional cash transfer programs in the 2000s reduced teenage pregnancy rates in Colombia, especially when the benefits were conditional on school success. For the case of Nigeria, Osili and Long (2008) estimate a reduction in fertility of women younger than 25 after the introduction of universal primary education.

As for the changes in quantity, women's education can translate into lower infant and child mortality rates. This can affect the number of births and reduce fertility (Kravdal, 2002; Baudin et al., 2015). Similarly, the introduction of modern contraceptive methods can reduce the risk of unwanted pregnancies, reducing also completed fertility.

Other factors to consider are the changes in the opportunity cost for women. An increase in women's education can create new labour market opportunities and raise the economic aspirations of women, therefore increasing the opportunity cost of childbearing (Easterlin & Crimmins, 1985; Castro Martin, 1995; G. Becker, 1960). Education can increase consumption aspirations which increases the relative cost of children.

Indirectly, education can spread information about family planning or new norms regarding family size and structure which can translate into more acceptance and use of contraceptive methods to limit fertility (Caldwell, 1980; Axinn & Barber, 2001).

4.2.3 Community-level effects: education and urbanisation

Finally, education attainment is also an aggregate phenomenon that can affect fertility levels in a specific community (Frye & Lopus, 2018). Education has the potential to create significant spillovers from the educated to the non-educated group by propagating new norms and behaviours (Frye & Lopus, 2018; Axinn & Barber, 2001). For example, Colleran, Jasienska, Nenko, Galbarczyk and Mace (2014) show that in the case of Poland, less-educated women adopted the low-fertility behaviour of the more educated women, which led to a convergence in fertility rates despite differences in educational levels. The transmission of the new norms and behaviours could be direct, by communication and interaction with peers or indirect through observation and imitation. Information about the use and availability of contraception could be passed from the more educated women to the uneducated while less-educated women can adopt preferences for a smaller number of children if they live in a community with a higher proportion of smaller families. Additionally, general increases in female education can transform social norms breaking old ideas and traditional modes of reproduction (Kravdal, 2002; Tienda, Diaz & Smith, 1985). Therefore, community-level characteristics could be important in fertility decisions, especially in a context of rapid educational change. As Tienda et al. (1985) argues in the case of Peru, once the majority of women in a cohort reached at least primary education, the educational composition of a community had a significant effect on the reproductive behaviour of women.

4.3 Education in Colombia: 1905-1970

Colombia represents a context in which the expansion of education happened fairly rapidly and this change coincided with other socioeconomic changes like the fertility decline. In this section, I briefly summarise the history of education in Colombia

during the 20th century.[4]

The country started the 20th century as one of the countries in Latin America with worse levels of literacy, but liberal governments of the 1930s reformed the educational system, expanding the right to access education and promoting universal education. After the 1950s educational levels started increasing significantly. As we will discuss in the next chapter, access to education continues to be an important issue in Colombia and universal education has not been achievable. Coverage rates in primary education are above 85% and in secondary education are close to 75%.

4.3.1 1905-1957

In 1900, 66% of the population of Colombia did not know how to read and write, and only 3.5% of the population was enrolled in primary schooling. By 1905 6.7% of the total population was enrolled in primary schooling in Antioquia and the number was as low as 1.2% in Atlantico (Ramírez-Giraldo & Téllez-Corredor, 2006). The more industrial regions in the Andean macro-region were able to invest more in education in comparison to other departments in the Caribbean.[5]

A similar gap emerged between rural and urban schools, and it was reinforced by differences in the type of schools that the Government designated for each area. Primary school in urban areas consisted of 6 years while in rural areas it consisted of only 3 years. Access to secondary education was limited and the number of enrolled students in secondary education during the first half of the 20th century was low because it was targeted at the educated and wealthy elites (Helg, 2001; Ramírez-Giraldo & Téllez-Corredor, 2006). Only 0.5% of the total population was enrolled in secondary schooling at the beginning of the 20th century and the proportion stayed

[4]See Fuentes-Vásquez (2021) for a history of education in the first half of the 20th century and Ramirez-Zuluaga (2016) for the second half of the 20th century. See also Helg (2001).

[5]The Andean region had on average higher enrolment rates in comparison to other macro-regions during the 20th century (Ramírez-Giraldo & Téllez-Corredor, 2006).

the same until the 1950s.

In terms of gender, women had lower enrolment rates than men in primary schooling and unequal access to education until the mid-1930s when the liberal party won the presidential elections for the first time in the century and started a series of reforms placing education at the centre of the Government efforts (Helg, 2001). The liberal Governments of the 1930s and 1940s improved gender equality with two important reforms. First, under the presidency of Olaya Herrera in 1933, the Government equalised the conditions and degree requirements between the schools of men and women in secondary education which in practice granted, for the first time, the possibility for women to aspire to a high school diploma (Iregui-Bohórquez, Melo-Becerra, Ramírez-Giraldo & Tribín-Uribe, 2020). During the presidency of Lopez Pumarejo in 1936, the Government mandated that primary education had to be compulsory and free, and legally banned discrimination against students for the reason of race, religion, social class or illegitimacy. But most importantly, this reform recognised the right of women to enter university on equal terms with men. In the mid-1940s, with the idea of promoting higher education for women, the Government established the "Colegio Mayores", which were institutions that offered university degrees without completing secondary studies. These reforms started the path towards mass education and helped to close the gender gap in education by the late 1970s (Ramírez-Giraldo & Téllez-Corredor, 2006; Fuentes-Vásquez, 2019).

During most of the 20th century, the costs of running schools were divided between the nation, the departments and the municipalities. However, these costs fell mainly on the departments, especially the cost of providing primary education. Until the 1940s, the central government was in charge of secondary education, although this was virtually limited to men. As expected, and despite the government's interest in education through the 1930s and 1940s, this division of the costs implied that the

richer departments were able to finance more education while poor departments lagged (Fuentes-Vásquez, 2019). Despite the efforts of the liberal governments, including the establishment of a Municipal Development Fund to promote mass education, the regional differences in education persisted, not only in coverage but also in the quality of the education (Fuentes-Vásquez, 2019).

Despite the liberal reforms, no significant changes occurred during this period mainly because of the lack of resources. It is only after the 1950s that the country saw a significant expansion in education (Ramírez-Giraldo & Téllez-Corredor, 2006; Helg, 2001).

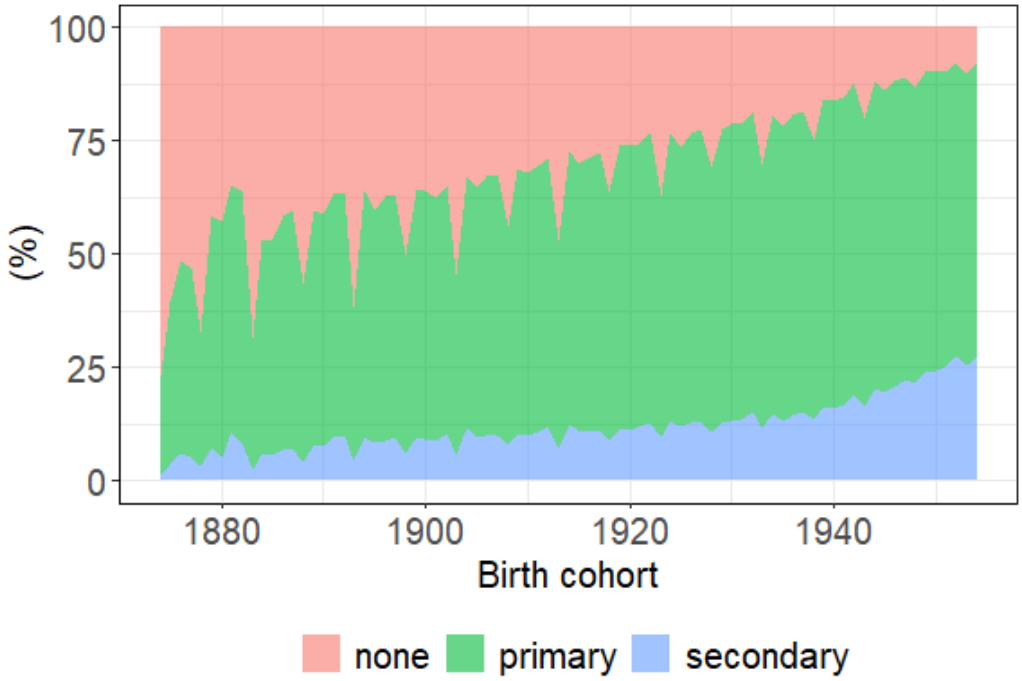
4.3.2 1958-1970

During the first 50 years of the 20^{th} century, the number of students enrolled in primary schooling steadily increase while the students in secondary schooling saw almost no change. But in the second half of the century, the expansion of education finally started and between 1950 and 1970 the century witnessed the sharpest increase in the number of students at all levels, but especially in primary schooling. As Ramírez-Giraldo and Téllez-Corredor (2006) argue, these changes were the result of the greater fiscal capacity that resulted from economic growth after several coffee booms. According to the 1973 census, more than 50% of the people born after 1900 had at least completed primary school. Women born in the 1910s were the first cohort of women to have more women with primary schooling completed than women with no schooling, while more than 35% of women born in the mid-1950s had completed secondary education as shown in Fig. 4.1.

The sustained increase in education after the mid-1950s in Colombia was a result of the plebiscite of 1957 (Ramírez-Giraldo & Téllez-Corredor, 2006). The government started a national plan to expand educational coverage and agreed to

spend no less than 10% of its budget on public education. The period saw an increase in literacy and enrolment rates as well as in teachers and educational establishments across the territory, in particular in primary schooling. The rate of increase in the number of students during this period was even higher than the rate of population growth and the increase in enrolment rates was particularly significant for women. By 1970 the proportion of women in education was 49% of the total (Ramírez-Giraldo & Téllez-Corredor, 2006). Nevertheless, women still lagged behind men, especially at the university level. In 1965, females represented no more than 20% of total university enrolment (Iregui-Bohórquez et al., 2020).

Figure 4.1: Women's education by cohort, 1875-1950



Notes: The graph shows the proportion of women in each educational category by the year of birth. The data comes from the Census of 1973 and is based on women older than 18 years old.

Although the public spending on education increased for the entire country, the increase did not occur in the same proportion at the sub-national level. Similar to what had happened during the first half of the 20th century, some departments were able to allocate more funds for education than others, and this translated into

differences in enrolment rates and the quality of education at the departmental level. During the 1950s, more than 11% of the total population was enrolled in primary schooling in Antioquia while in Atlantico or Bolivar this proportion was less than 6.5% (Ramírez-Giraldo & Téllez-Corredor, 2006). These differences increased over time creating an educational gap that persists until today. According to the Minister of Education, in 2017, the coverage rate in secondary education in Bogota was 90% while in Amazonas, Arauca and Vaupes were less than 10%.

4.4 The long-run relationship between education and fertility

As fertility and education experience similar changes around the same time, the relationship between fertility and education has been broadly studied in Colombia. Using a 4% census sample of the 1973 census, (Potter et al., 1976) shows that the gap in fertility levels between educated and non-educated women remained unchanged during the fertility transition. By 1990, Heaton and Forste (1998) find that the gap between uneducated and educated women was closing as declines in fertility were larger for the least educated women. However, in a recent paper, Batyra (2020) challenges this finding and shows that in 2005 the differences in rates of first birth between the lowest and highest educated women were still evident in Colombia. Given that these studies analyse different fertility measures it is difficult to draw a clear pattern of the long-run relationship between education and fertility.

This section shows new stylised facts about the long-run relationship between female education and fertility at the national, sub-national and individual-level using individual-level census data. As in my two previous chapters, my empirical work is based on the complete 1973 census, and I also use the 10% samples of the 1985 and

1993 censuses from IPUMS (Minnesota Population Center, 2019).[6]

4.4.1 National

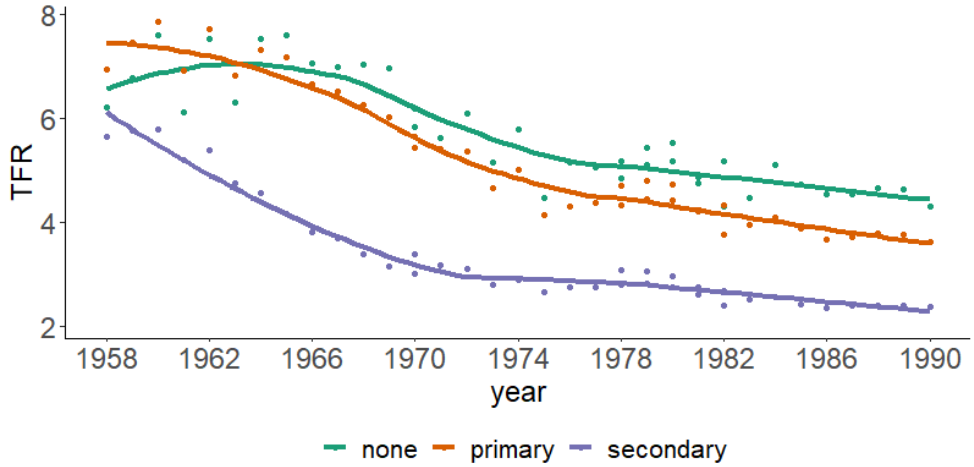
Total Fertility Rate

I estimate Total Fertility Rates between 1958 and 1990 by the educational category of the mother. The estimations are done following the Own Child Method described in chapter 1. Fig. 4.2 shows that fertility declined continuously until 1990 in all educational groups, but it shows that fertility levels were already descending in 1958 for women with secondary and probably primary education.[7] Fertility started falling around 1962 for women with no education. From 1962 to 1970, fertility dropped in all schooling categories but it declined faster in secondary schooling, amplifying the gap between educated and non-educated women. Between 1970 and 1980 the fall in the total fertility rate slowed down for the highest educational category, while accelerated for women who had completed primary schooling. From 1980 to 1990 fertility kept declining at a slower rate in the three educational categories, but the gap in fertility levels between highly-educated and non-educated women remained almost unchanged. During the fertility decline, fertility dropped in all educational categories but women with secondary schooling had a faster decline from 6 to 2 in 30 years, whereas for women with no education fertility dropped from 8 in 1960 to 4 in 1990.

[6]Chapter 1 describes these censuses in detail.

[7]The estimations for women with no education fluctuate more, probably due to age-heaping.

Figure 4.2: Total Fertility Rates by educational levels of mothers



Note: Calculations implementing the OCM. Women between 15 to 78 years old and children between 0 to 14 years old. See Section 2.6 for a description of the methodology. Sources: Authors' calculations based on 1973 full Census data, and 1985 and 1993 Census samples from IPUMS-International.

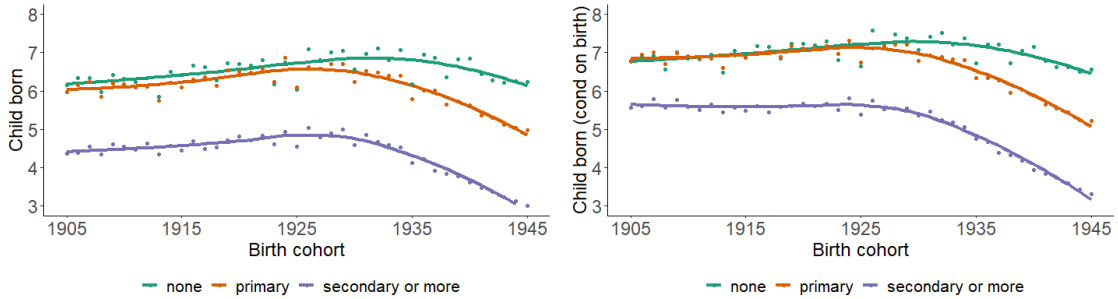
Intensive margin

To distinguish the effects of education on the intensive and the extensive margin, I estimate the number of children ever born to women ages 48 to 68 by educational level.[8] As with the TFR, the results of completed fertility in Fig. 4.3 reveal a clear educational gradient in fertility even before fertility started declining. This is particularly noticeable for women with primary or less than primary education and women with secondary education or more. As shown by Fig. 4.3, the gap seems to amplify after the 1930 cohort, when fertility declines for all educational groups. If we compare the two graphs, the fertility decline was particularly rapid for mothers with secondary education born after 1930 (see Fig. 4.3b). While fertility declined across all groups, the fertility gap increased.

[8]I select this age range because it allows me to observe fertility patterns in the long-run. Furthermore, this age group allows me to observe women with completed fertility in the 3 censuses. To avoid overlapping in the information of the censuses the information for 1905 to 1923 comes from the 1973 census, the information for 1924 to 1935 comes from the 1985 census and the information for 1936 to 1945 comes from the 1993 census.

Figure 4.3: Intensive margin: children ever born

(a) Children born (b) Child born cond. on having a birth



Note: The graphs show the total number of children ever born to women between 48 to 68 years old using the reported variable in the censuses. The information for 1905 to 1923 comes from the 1973 census, the information for 1924 to 1935 comes from the 1985 census and the information for 1936 to 1945 comes from the 1993 census. Panel a presents the results for all women and Panel b the results for all women (conditionally on being mothers). Sources: Authors' calculations based on 1973 full Census data, and 1985 and 1993 Census samples from IPUMS-International.

Extensive margin

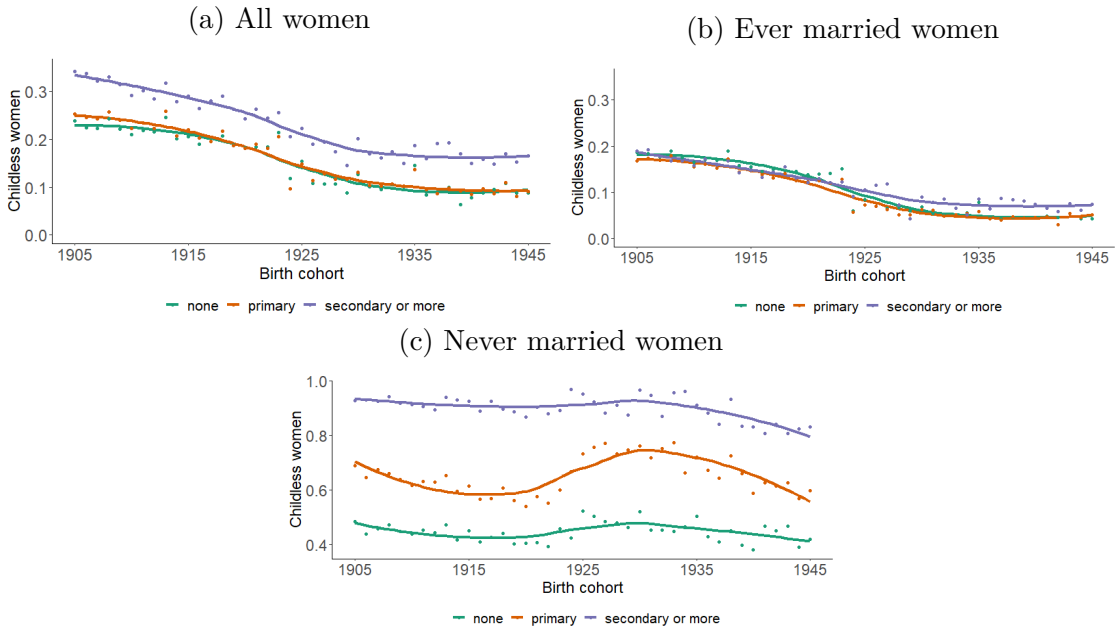
Then, I estimate the percentage of women that remained childless for the same cohort of women (born between 1905-1945). I also estimate this for ever-married women (which includes women in a partnership, divorced and widows) and non-married (which includes only women that declare to be single). Looking at all women, the results show that in comparison to fertility rates and completed fertility, the educational gradient vanishes while childlessness declines steadily. The fall in childlessness is similar to the estimations of Baudin et al. (2015) for the US case where childlessness falls from 20% to around 10% from 1905 to 1945 (see Fig. 4.4).

The simultaneity of the decline of childlessness within marriage across all education groups indicates that, as shown by Aaronson et al. (2014) for the US case, the share of couples that did not have children declined as schooling opportunities expanded after the 1930s. These results suggest that voluntary childlessness was the main driver of childlessness for couples. The absence of an educational gradient suggests that sexual abstinence, and sterility due to venereal diseases did not vary across educational levels as much as they did in pre-industrial England (De La Croix

et al., 2019). These findings indicate that changes in education, health and economic conditions could partially explain the drop in childlessness within marriage (Castro & Tapia, 2021).

The results for single women confirm that the fertility decline was accompanied by a fall in childlessness in ever-married women and not a fall in childlessness in single women. In fact, childlessness in single women remained more or less constant across this period, and with a clear educational gradient. The results show that non-educated single women were significantly more likely to have a child than single women with secondary education. Almost 90% of single women with secondary education remained childless while more than half of the non-educated women were mothers.

Figure 4.4: Extensive margin: proportion of childless women



Note: The graphs show the percentage of women between 50 to 63 years that remained childless. Panel a presents the results for all women, Panel b the results for ever-married and Panel c for unmarried women. Ever-married women include women that report being: married, cohabiting, divorced or widows. Unmarried women are women that report being single. Sources: Authors' calculations based on 1973 full Census data, and 1985 and 1993 Census samples from IPUMS-International.

All in all, women born after the 1930s had fewer children, and a smaller

Table 4.1: Wanted children vs. children ever born

N. of children	0 yrs of schooling	1-6 yrs of schooling	7+ yrs of schooling
1976	7.4	4.9	2.8
1986	4.8	3.9	2.3
N. of wanted children			
1976	3.9	2.7	2.2
1986	2	2.2	1.8

Notes: The information for 1976 comes from the WFS, while the information for 1986 comes from the DHS. Source: From Bongaarts and Lightbourne (p.235-236:1996) in Guzman et.al (1996)

but constant proportion of these women remained childless. This effect is bigger for women with secondary education.

It should be noticed that although fertility declined across all educational groups from 1950 to 1990, I do not observe convergence in fertility levels, either when I look at the TFR or using the number of children ever born. A similar finding was discussed in Bongaarts and Lightbourne (1996) when looking at the number of children born and the number of wanted children in Colombia between 1976 and 1986. As shown in Table 4.1, fertility differentials by levels of education were substantive but diminished over time. As for fertility preferences, they find relative convergence in the number of children wanted between 1976 and 1986. Variation in fertility by 1986 was explained by variation in unwanted fertility rather than in fertility preferences.

It is unclear what allowed convergence in fertility preferences but education has the potential to create significant spillovers from the educated to the non-educated group propagating new norms and behaviours (Frye & Lopus, 2018; Kravdal, 2002; Axinn & Barber, 2001). In that scenario, fertility levels will converge despite divergences in educational outcomes (Colleran et al., 2014). Education can also spread information about family planning or new norms regarding family size and structure which can translate into more acceptance and use of contraceptive methods to limit fertility which can explain the variation in the number of unwanted children

(Caldwell, 1980; Axinn & Barber, 2001). However, understanding these mechanisms requires more disaggregated data. Sub-national data can reveal if the expansion of education varied at the departmental level due to the financial capabilities and the political will of each department.

4.4.2 Departmental

To examine if departmental differences in educational coverage were associated with fertility levels, I study the association between schooling rates and fertility rates in 1960 and 10 years later. During these years, female education saw steady growth in most departments and national territories (Fuentes-Vásquez, 2019). As for the changes in the TFR, and in line with the discussion of Chapter 2, all departments and national territories experienced a fertility decline. However, one might like to understand whether departments that expanded education more rapidly also experience greater fertility decline.

Fig. 4.5 displays the association between the absolute change in the total fertility rate between 1960 and 1970 at the departmental level and the absolute change in female education, looking at primary and secondary education separately.[9] The results show that while fertility declined in all departments, female primary and secondary education grew across the whole country, although departments followed different paths.

Fig. 4.5a shows that changes in female primary schooling were negatively associated with changes in fertility, which means that places that saw a faster expansion of education, experienced a slower fertility decline. The coefficient for the independent variable, -0.051, represents the average change in the total fertility rate for one standard deviation change in education. For example, if education increased by 1

[9]To ease the interpretation changes in education are standardised using a Z-score method.

standard deviation, we expect the total fertility rate to decrease by 0.018. However, the low R^2 (of 0.008) indicates that the model does not explain a large amount of the variance in the dependent variable. This suggests that there may be other important factors that are not being taken into account in the model. This weak relationship can also be seen by comparing individual departments. For example, Caqueta and Sucre saw faster fertility declines in comparison to Cauca and Amazonas despite experiencing a similar change in primary schooling rates. Even more, the intercept, -2.39, represents the change in the total fertility rate when the change in education is 0 standard deviations. This implies that even places that did not experience changes in enrolment in primary education saw a decline in fertility rates.

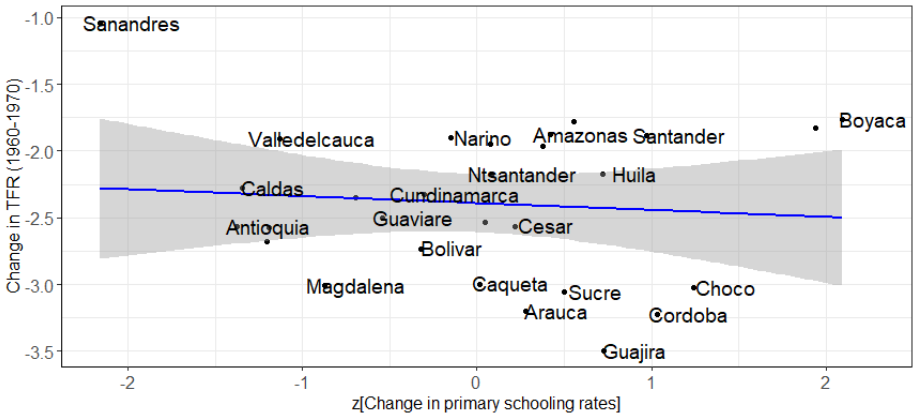
The patterns are somehow different for secondary schooling. Fig. 4.5b shows that the relationship between fertility change and expansion of secondary schooling is positive, meaning that one standard deviation in changes in secondary education would increase the total fertility rate by 0.089. However, as in the case of primary education, the model only explains a small fraction of the variance in the total fertility rate, as indicated by the low R^2 value. Again, comparing departments that had a similar expansion in education confirms this finding. Take the departments of Cordoba and Cauca. Both experienced a similar change in secondary schooling rates, but their changes in fertility rates were significantly different. Caqueta saw a drop of around 3 children per woman between 1960 and 1970 while Cauca saw a drop of less than 2. Similarly, the intercept reveals that places that did not experience any change in enrolment rates in education saw a decline in fertility rates of 2.39 children per woman.

Although female education expanded, at the departmental level, fertility changes were weakly associated with educational changes. The results confirm that the fertility decline in Colombia cannot be explained only through the expansion of

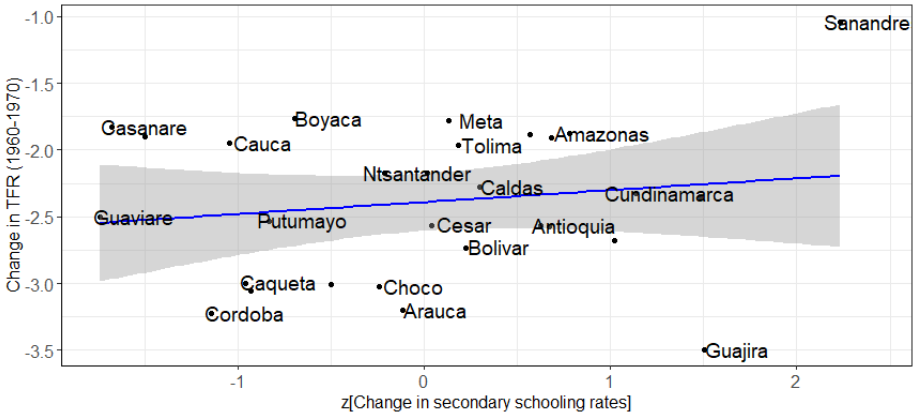
female education or because more women had secondary schooling (also known as compositional effect).

Figure 4.5: Changes in fertility and education rates

(a) Primary



(b) Secondary



Notes: Panel a shows the relationship between the absolute change in the total fertility rate between 1960 and 1970 at the departmental level and the absolute change in female primary education. Panel b shows the same relationship but with secondary education. The blue line is a linear fitted line. For panel a the predicted equation is $y = -2.39 - 0.051x$ with an R-squared of 0.008. For panel b the predicted equation is $y = -2.39 + 0.089x$ with an R-squared of 0.0251. Sources: The departmental TFR is estimated following chapter 1. Female schooling is estimated as the share of women age 15 to 45 that have primary or secondary education using the 1973 census.

4.4.3 Individual

To fully disentangle the effects of women's education on fertility, I exploit the censuses' individual-level character and incorporate individual and community-level variables to examine the relationship between fertility decisions and women's education. I look

at both the intensive and the extensive margin of fertility for women with completed fertility in the census of 1973 and the samples for 1985 and 1993. Additionally, I study the prevalence of starting, spacing and stopping in the census of 1973, when fertility was changing more rapidly. The model is shown in Eq. (4.1).

$$\begin{aligned} Y_{i,a,m} =& \beta_0 + \beta_1 * \text{education level}_i + \beta_2 * \text{Z}[\%\text{ women secondary education}]_{a,m} + \\ & \beta_3 * \text{Z}[\text{urbanisation rate}]_m + \mathbf{X}_i\lambda + \epsilon_{i,a,m} \end{aligned} \tag{4.1}$$

Where $Y_{i,a,m}$ corresponds to a fertility measure or behaviour for a woman i, age a and living in a municipality m. The coefficient β_1 is the estimated effect of the individual education of a woman on fertility, β_2 is the estimated effect of the percentage of women with secondary education for each cohort of women at the municipal level and β_3 is the estimated effect of the level of urbanisation. Both community variables are standardised using a *Z-score* method to ease interpretation. $\mathbf{X}_i$ is a vector of individual controls such as age, urban location, access to public services, socioeconomic status and marital status. Other unobserved factors are captured with the random error term $\epsilon_{i,a,m}$. The standard errors are robust and clustered at the municipal level.

Table 4.2 shows the estimated effects of schooling and community variables on the fertility of women with completed fertility born between 1910 to 1940. The fertility outcome in the first three columns is the number of children ever born to women older than 53 years old conditional on having a birth (intensive margin). In columns 4 to 6 the fertility outcome is an indicator variable equal to zero if the woman has at least one child and one if the woman is childless (extensive margin) and the coefficients are presented as odd-ratios, to make the interpretation of the effects easier.

For women born between 1910 and 1920 primary education increased the total number of children in comparison to women with no education, although it increased the probability of remaining childless by almost 3%. Secondary education reduced the total number of children and further increased the probability of remaining childless by 46%. For the women born between 1922 and 1940, the educational gradient is clear: both primary and secondary education reduce the total number of children and increases the probability of remaining childless. The effects are similar across the cohorts, although education seems to have a greater effect in the 1933 - 1940 cohort, where secondary education reduces the total number of children by more than 1 child and increases the probability of remaining childless by 124%.

In the three cohorts, the effects are economically bigger for secondary education than for primary education. This relationship can emerge because more educated women face a higher opportunity cost of childbearing and can achieve greater relative autonomy in comparison to less-educated women (Jejeebhoy, 1995).

Peer effects, such as the percentage of women with secondary education in the woman's cohort in her municipality have a negative effect on the number of children for the 1922 to 1940 cohort, while they are absent in the first cohort. For women born after 1922, an increase of one standard deviation in the percentage of women with secondary education translates into a reduction of the fertility of around 0.28. Similarly, the effects of the urbanisation rate of her municipality are moderate and an increase of one standard deviation in the urbanisation rate decreases fertility by 01.4. For the extensive margin, the changes in urbanisation rate reduce the probability of remaining childless in the 1910-1920 cohort while the percentage of women with secondary education increased the probability of remaining childless for the 1922-1932 cohort. Overall these results suggest that in the case of Colombia, education affected fertility decisions mainly directly while community variables have

a smaller effect.[10]

Table 4.2: Estimated effects of schooling and community variables on fertility

	Intensive: Num. of children			Extensive: Remaining childless		
	1910-1920	1922-1932	1933-1940	1910-1920	1922-1932	1933-1940
	(1)	(2)	(3)	(4)	(5)	(6)
Individual variables						
no schooling (reference)						
primary	0.2816***	-0.7074***	-0.8237***	1.0231*	1.4226***	1.4396***
	(0.0646)	(0.0497)	(0.0443)	(0.0130)	(0.0367)	(0.0454)
secondary	-0.7359***	-1.186***	-1.300***	1.4602***	1.9618***	2.2438***
	(0.0724)	(0.1452)	(0.0631)	(0.0322)	(0.0705)	(0.0561)
Community variables						
Z[% of women secondary in age-cohort]	-0.1654	-0.2821***	-0.2856***	1.0163	1.1774***	1.0128
	(0.1099)	(0.0894)	(0.0318)	(0.0196)	(0.0532)	(0.0283)
Z[urbanisation rate]	0.0401	-0.1436**	-0.1085*	0.9317***	0.9323	1.0212
	(0.1135)	(0.0718)	(0.0553)	(0.0248)	(0.0529)	(0.0549)
(Intercept)	9.428***	7.335***	2.583***	3.080***	1.487***	2.737***
	(0.2866)	(0.4453)	(0.5545)	(0.0884)	(0.3413)	(0.5501)
S.E Clustered	Municipality	Municipality	Municipality	Municipality	Municipality	Municipality
Observations	346,107	64,584	68,602	443,855	70,226	73,191
Adjusted R^2	0.04	0.03	0.10	0.01	0.24	0.25
Method	OLS	OLS	OLS	Logit	Logit	Logit

Notes: Columns 1 to 3 estimate the effect of education on the intensive margin (number of children ever born conditional on ever having a birth) and Columns 4 to 6 estimate the effect of education on the extensive margin and to ease the interpretation of the coefficients, these are reported as odd-ratios. The community variables percentage of women with secondary education by age cohort and the urbanisation rate are standardised using a *Z-score* method. The samples consider only women with completed fertility. The cohort of women born between 1910-1920 comes from the 1973 census, the cohort born between 1922-1932 comes from the 1985 census and the cohort born between 1933-1940 comes from the 1993 census. All models include individual controls such as age, urban location, access to public services, socioeconomic status and marital status. Standard errors in parentheses are clustered at the municipality level, and $***$: 0.01, $**$: 0.05, $*$: 0.1 indicate statistical significance. For Logit models I report the squared correlation instead of R^2.

I then turn my attention to other fertility decisions, such as the age of starting, spacing between births and stopping. Starting is defined as the age at which a woman starts her maternal life, and commonly an earlier starting age would translate into higher fertility. In this chapter, I proxy starting with the probability of having a child when a woman is between 19 to 22 years old. This age range allows me to study childbearing at an early age while at the same time focusing on women that should have already completed secondary schooling to alleviate the concern of simultaneity in the decision-making between childbearing and schooling. Spacing is defined as the average of all intervals between childbirths and I estimate this for women who had more than one child and are aged 23 to 36. As child mortality and leaving home can affect the measurement of spacing and these issues are more problematic when linking older mothers to older children, I do the analysis of spacing

[10]Table 4.4 reproduces the estimations using municipality-fixed effects instead of community variables. The results show that individual-level effects are robust to the inclusion of these fixed effects and the coefficients change marginally.

for women between 23 and 25 separate from women between 26 to 36, to minimise this concern.[11] Although longer spaces could reduce the total number of live births, when stopping is achievable earlier through contraception, a shorter spacing could imply that a woman focuses her fertility efforts in a narrower period. Stopping corresponds to the age at which a woman stops having children (usually once she attained her desired number of children) and it is proxied in this chapter with the probability of having a child at 40 to 45.

The results are shown in Table 4.3. Column 1 shows the relationship between education and starting. In terms of prevalence, in 1973 around 37% of women started their fertility life at ages 19 to 22. The results reveal that there is a clear and significant educational gradient: obtaining primary schooling reduces the probability of having a birth at this age by 26% in comparison to women with no education but the effect is bigger when looking at secondary schooling (a reduction of 71%). On the contrary, the percentage of women with secondary education and the level of urbanisation of the municipality of residence are not associated with the probability of having a child at a young age. These results confirm that having access to secondary education had a key effect on fertility while there is no evidence of community-level effects. Additionally, these results are consistent with the idea that individual-level exposure to schooling directly affects the behaviours of students. The literature has found that students with improved literacy, numeracy, and cognitive skills wait longer to have sex (Grant & Hallman, 2008; Marteleto, Lam & Ranchhod, 2008). Also, education is related to an increase in aspirations for a career which as Jensen (2012) shows results in fewer children.

The results for spacing in Columns 2 and 3 show that the educational gradient also appears in this strategy and more educated women reduce their spacing in

[11]Unfortunately, the presence of twins can also create bias in the measurement of spacing and I cannot account for this with the census data.

comparison to less educated women. This result suggests that women with secondary education are concentrating their fertility efforts on a shorter period by reducing the space between births. This is the case for young women (23-25 years old) as well as for the older group (26-36 years old). Women with secondary education start later, have a shorter space between births and as Column 4 shows stop earlier in comparison to women with no education. As for the community-effects, the percentage of women with secondary education has no significant effect on the size of the space, while the urbanisation rate of the municipality moderately reduces the average size of spacing.

The results in Column 3 indicate that on average, 82% of women stopped before the age of 40 and the probability of having a birth at ages 40 to 45 was mainly affected by having secondary education. Similar to the results of completed fertility, the majority of this cohort of women, born between 1928 and 1933, had achieved primary education while around 15% finished secondary school. Therefore, the more disadvantaged women that did not have access to any education were more likely to have birth at these ages, and their probability of childbearing was not affected by the educational level of their peers. This result also suggests that for this cohort of women, education could be reflecting individual-level social status or income, which are important confounders and limit the interpretation of this result.[12] Nonetheless, the results rule out potential community effects and demonstrate that by 1973 the educational gradient existed at the individual level and not across municipalities. The results are consistent with the results of Table 4.2 and this finding implies that economic development that happens at the municipal level, like urbanisation and changes in education, cannot fully account for the fertility decline.

[12]Although all the specifications include a set of controls for socioeconomic status with the information available in the census to minimise this issue.

Table 4.3: Estimated effects of schooling and community variables on limiting behaviours in 1973

	Starting: Any birth age 19-22 (1)	Spacing: Interval between births ages 23-25 (2)	Spacing: Interval between births ages 26-36 (3)	Stopping: One more birth age 40-45 (4)
Individual variables				
no schooling (reference)				
primary	0.7410***	-0.1879***	-0.1792***	1.0285
	(0.018)	(0.0090)	(0.005)	(0.018)
secondary	0.2888***	-0.2805***	-0.3439***	0.5685***
	(0.030)	0.0199	(0.021)	(0.038)
Community variables				
Z[% of women secondary in age-cohort]	0.9617	0.0115	0.0026	0.9862
	(0.0375)	0.0140	(0.0137)	(0.003)
Z[urbanisation rate]	0.9770	-0.0292**	-0.0247*	0.878***
	(0.0375)	0.0136	(0.0135)	(0.0008)
(Intercept)	-1.808***	3.349***	4.698***	-2.401***
	(0.2229)	(0.0251)	(0.0337)	(0.303)
S.E Clustered	Municipality	Municipality	Municipality	Municipality
Observations	755,111	182,746	782,039	456,047
Pseudo R^2	0.45	0.05	0.13	0.10
Method	Logit	OLS	OLS	Logit

Notes: Column 1 estimates the probability of having a child for women between 19 to 22 years old using a logit model and to ease the interpretation of the coefficients, these are reported as odd-ratios. Columns 2 and 3 estimate the mean size of the interval between 2 children for women that are between 23 and 36 years old and the estimations are done using an OLS model. Column 4 estimates the probability of having one more birth for women with at least one child that are between 40 to 45 years old and the estimations are done using a logit model. The community variables percentage of women with secondary education by age cohort and the urbanisation rate are standardised using a *Z-score* method. As for previous Logit models, to ease the interpretation of the coefficients, these are reported as odd ratios. All models include individual controls such as age, urban location, access to public services, socioeconomic status and marital status. Spacing also includes the number of children as a control. Standard errors in parentheses are clustered at the municipality level, and $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ indicate statistical significance.

The results of this section support the existence of a link between women's education and fertility and indicate that the gap between educated and uneducated women in fertility existed during the fertility decline and that uneducated women were unlikely to adopt simultaneously the fertility behaviour of more educated women, although this does not rule out that uneducated women could adopt this behaviour with a lag. The results also allow us to conclude that access to secondary education was a key factor behind the differences in fertility. Additionally, the results suggest that women with the same educational status will display similar fertility behaviour across all cohorts.

The results of a small effect of the community variables make sense in a society with little interaction across different classes, in which the process of learning from the peer is limited by the places in which peers interact. This, especially the result on the percentage of women with schooling in the municipality, make sense in a setting like Colombia where schooling has been segregated and where there is little interaction across social classes, as we will see in the next chapter. However,

a final caveat is that the municipality could also not be the right level to measure community effects as municipalities vary in population size and socioeconomic development. Tienda et al. (1985) argues in the case of Peru, once the majority of women in a cohort reached at least primary education, the educational composition of a community had a significant effect on the reproductive behaviour of women. This could be the case in a smaller level of aggregation, such as neighbourhoods.

4.5 Conclusion

This chapter provided new stylised facts of the long-run relationship between fertility and women's education. I focused on the period after the liberal reforms of the 1930s that granted access to secondary education to females and promoted mass education across the country. Using individual-level census data I estimated the relationship between education and fertility at several levels of aggregation and along the two margins of fertility. The findings caution that the relationship between fertility and women's education is not necessarily monotonic.

At the national level, the fertility gap between the lowest and highest educated women existed during the fertility transition and the gap did not close by 1990. However, fertility declined across all educational groups. The results at the departmental level confirm this. Fertility decline happened broadly everywhere and changes in fertility rates were mostly independent of changes in female schooling rates, both in primary and secondary schooling. These findings suggest that the fertility decline, at the aggregated level, cannot be explained through changes in education.

At the individual level, the relationship between schooling and fertility holds strongly. For most cohorts, both primary and secondary education increased the probability of remaining childless, reduced the total number of children and reduced the probability of having a birth at a younger age. It is important to highlight

that in all cases, the effect of secondary schooling is significantly bigger, which poses problems of interpretation, as secondary schooling can also be related to social status and income.

Peer effects, such as the percentage of peers with secondary education, are mostly ruled out as these had a small effect on the total number of children, the probability of remaining childless and the use of strategies to limit fertility (e.g. starting, stopping). As fertility declines smoothly also for non-educated women, it is relevant to further investigate what explains the decline in this group of women, given that educational peer effects seem to matter little.

Finally, the results of the extensive margin in single women deserve attention. For this group of women, the educational gradient is clear and stable during the period studied. Single women with secondary education remained childless in a much higher proportion (up to 90%) than non-educated women (close to 50%), and the trend changes little over time. Although the causes of this pattern are not explored in the chapter, it is clear that single women without education and with children represent one of the most vulnerable population groups, as uneducated single mothers generally have less earning power and their children are likely to have worse socioeconomic outcomes. Understanding why the educational gradient exists for single women and not for married women is crucial to disentangle the mechanisms through which education affects fertility.

4.6 Future research on women's fertility and education

The new stylised facts of the long-run relationship between fertility and women's education that are presented in this chapter suggest that the fertility transition, at

the national and sub-national levels, cannot be explained only through changes in education. But the findings at the individual level show that access to education is negatively and strongly associated with fertility in both the extensive and intensive margins. In particular, women born after 1922 exhibit a clear educational gradient that does not appear for women born before 1920. The strongest effect is the effect of secondary schooling

In this chapter, I did not address the causal link between women's education and fertility but in this section, I outline two potential research paths to understanding better how education, and which type of education, could have shaped fertility decline.

The first research avenue is the evaluation of the two most important educational reforms in the liberal period of the 1930s. Under the presidency of Olaya Herrera in 1933, the government equalised the conditions and degree requirements between the schools of men and women in secondary education which in practice granted, for the first time, the possibility for women to aspire to a high school diploma (Iregui-Bohórquez et al., 2020). Although this first reform did not change the price of education, it did modify the incentives to complete secondary schooling. If prolonged exposure to schooling affects fertility behaviour, for example by delaying the age at first birth, we should expect that women that were exposed to this reform end up with fewer children than women that were not exposed. Three years later, the Government mandated that primary education had to be compulsory and free and recognised the right of women to enter university on equal terms with men. This reform affected the price of primary education for both men and women while opening the door to university education for women. As these reforms were effective at a particular date, it is possible to exploit the sharp discontinuities in who was affected by these reforms and I can compare the fertility of those women that were born before and after the introduction of the laws using a Regression Discontinuity

Design (RDD). The data from both year of birth, years of schooling and fertility are available in the census data.

A second avenue is to explore the effects of a radio-schooling programme that started in 1947 with the creation of Radio Sutatenza on fertility. The station, which was the first Catholic radio station in Latin America, was originally founded in the town of Sutatenza in the department of Boyaca in 1947 with the idea of combating adult illiteracy. By 1957, there were 855 parishes with a radio school and 136,385 persons "registered": of this, 30% were illiterate women.[13] The targeted audience of the radio station was illiterate peasants living in isolated, dispersed, and impoverished parishes. By the early 1960s Radio Sutatenza and its parent organization, Acción Cultural Popular (ACPO), worked with the Government to spread "scientifically based health and sex education materials and promote acceptance of modern approaches to sanitation, cooperative association, and technology" (Roldán, 2014). As I mentioned in the previous chapter, Radio Sutatenza also collaborated with Profamilia in 1973, when they launched the national campaign "Responsible Procreation". This non-formal education through radio could have changed the attitudes and social norms regarding fertility decisions and preferences and it could have increased the economic aspirations of parents, in particular of non-educated rural women.

4.A Appendices

As a robustness check for the individual level effects, I replace the community variables for municipal fixed effects, as municipal fixed effects are a better control for potential omitted variables. Community variables can be related to other economic variables that were not included in the original regression while municipal fixed effects control for both observed and unobserved municipality characteristics. The results confirm

[13]The annual report of official statics has statistics on the number of registered students from 1954 to 1965. In the report, this information is collected under the category "basic education for adults".

that the effects of individual education on fertility are not driven by potential omitted variables and are robust to the inclusion of municipal fixed effects.

4.A.1 Robustness: Municipal fixed effects

Table 4.4: Effects of schooling with municipal fixed effects - fertility

	Intensive: Num. of children			Extensive: Remaining childless		
	1910-1920	1922-1932	1933-1940	1910-1920	1922-1932	1933-1940
	(1)	(2)	(3)	(4)	(5)	(6)
no schooling (reference)						
primary	0.0663	-0.6775***	-0.7850***	0.9877	1.3513***	1.3903***
	(0.0663)	(0.0530)	(0.0403)	(0.0111)	(0.0316)	(0.0388)
secondary	-0.8378***	-1.179***	-1.280***	1.4260***	1.9246***	2.1789***
	(0.0586)	(0.1535)	(0.0648)	(0.0326)	(0.0693)	(0.0546)
Municipal f.e.	Yes	Yes	Yes	Yes	Yes	Yes
S.E Clustered	Municipality	Municipality	Municipality	Municipality	Municipality	Municipality
Observations	346,107	64,584	68,543	443,855	70,226	73,132
R^2	0.08	0.07	0.13	0.03	0.27	0.25
Method	OLS	OLS	OLS	Logit	Logit	Logit

Notes: Columns 1 to 3 estimate the effect of education on the intensive margin (number of children ever born conditional on ever having a birth) and Columns 4 to 6 estimate the effect of education on the extensive margin and to ease the interpretation of the coefficients, these are reported as odd-ratios. The samples consider only women with completed fertility. The cohort of women born between 1910-1920 comes from the 1973 census, the cohort born between 1933-1940 comes from the 1985 census and the cohort born between 1933-1940 comes from the 1933 census. All models include individual controls such as age, urban location, access to public services, socio-economic status and marital status and also municipal fixed effects. Standard errors in parentheses are clustered at the municipality level, and $***$: 0.01, $**$: 0.05, $*$: 0.1 indicate statistical significance. For Logit models I report the squared correlation instead of R^2.

Table 4.5: Effects of schooling with municipal fixed effects - limiting

	Starting: Any birth age 19-22	Spacing: Interval between births ages 23-38	Stopping: One more birth age 40-45
	(1)	(2)	(3)
no schooling (reference)			
primary	0.7183***	-0.1495***	1.0141
	(0.016)	(0.009)	(0.019)
secondary	0.2806***	-0.1218***	0.5745***
	(0.031)	(0.019)	(0.04)
Municipal f.e.	Yes	Yes	Yes
S.E Clustered	Municipality	Municipality	Municipality
Observations	755,111	1,162,194	456,066
Pseudo R^2	0.38	0.06	0.10
Method	Logit	OLS	Logit

Notes: Column 1 estimates the probability of having a child for women between 19 to 22 years old using a logit model and to ease the interpretation of the coefficients, these are reported as odd-ratios. Column 2 estimates the size of the interval between 2 children for women that are between 23 and 38 years old and that had at least 2 children and the estimations are done using an OLS model. Column 3 estimates the probability of having one more birth for women with at least one child that are between 40 to 45 years old and the estimations are done using a logit model and the results are also reported as odd-ratios. All models include individual controls such as age, urban location, access to public services, socio-economic status and marital status. Standard errors in parentheses are clustered at the municipality level, and $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$ indicate statistical significance.

Everybody knows the fight was fixed

The poor stay poor, the rich get rich

That's how it goes

Everybody knows.

Everybody Knows

Leonard Cohen

Part II

The long shadow: inequality of opportunity

5

Segregation in education: a long-term perspective

With Andrés Álvarez. Facultad de Economía - Universidad de Los Andes.

5.1 Introduction

Colombia is among the most unequal and stratified countries in the world. According to several contemporary studies, Colombia has one of the highest intergenerational income elasticity, one of the highest Gini coefficients worldwide and mobility in the country is lower in the upper and lower parts of the socioeconomic distribution.[1] The literature on inequality in Latin America, and in particular, in Colombia, has pointed to education as a key mechanism to close social gaps (Torche, 2014; Behrman, Gaviria & Szekely, 2001). However, several studies have indicated that the educational system in the region also reproduces patterns of social exclusion, which deepens the gap between social classes (Behrman, Gaviria & Székely, 2003; Marshall & Calderon, 2006).

A recent study by Cardenas, Fergusson and Garcia-Villegas (2021) underlines the presence of what they call an *educational apartheid* in Colombia, a form of social segregation in access to educational institutions. According to Fergusson and Flórez (2021a) while children from households in the lowest socioeconomic level represent more than 40% of students who complete secondary education, this group barely represents 10% of those who access tertiary education. On the other extreme, children from households of the highest socioeconomic level represent less than 5% of students who complete secondary education but represent more than 50% of those who access tertiary education.[2] Similarly, Duarte, Bos and Moreno (2012) show that in Colombia there are high inequalities in the academic results of students, associated with the socioeconomic level of their families and the type of school they attend (public vs. private). This gap is more important between schools than within them,

[1] In 2019 the World Bank reported a Gini index for Colombia of 51.3 and the OECD report of 2018 shows that in Colombia it would take at least 300 years for offspring of low-income families to reach the mean income. See also Narayan et al. (2018); OECD (2018); Angulo et al. (2012); Alvaredo and Londoño Velez (2013)

[2] See Figure 9 in Cardenas et al. (2021). The socioeconomic level corresponds to a stratification system used to target cross-subsidies for the payment of utilities in Colombia.

which denotes high segregation of Colombian schools according to the socioeconomic level of the students. Furthermore, the authors also show that access to school is the most visible and functional marker of social status that can be found in Colombia.

The goal of this chapter is to evaluate whether social segregation in education today has deep roots in the past. There are several reasons to believe that in Colombia social status has persisted since colonial times. First-pass evidence suggests that the rigid social structure of the Spanish colonial regime is reflected in contemporary Colombia, where Indigenous and Afro-Colombians appear at the bottom of the social ladder. Additionally, the literature has shown that the colonial regime set the type of institutions that we expect to be persistent (Acemoglu, García-Jimeno & Robinson, 2012; García-Jimeno, 2005).[3] Previous literature has shown that extractive institutions such as the *encomienda* and slavery had negative long-term effects at the national and sub-national levels, but the effects of these institutions on access to education have not been measured yet.

Due to the lack of reliable longitudinal data, it has been difficult to show if inequality and lack of social mobility have been persistent and stable over time or if, on the contrary, these processes can respond to public policies and institutional changes. However, the use of rare surnames and their relative representation in socioeconomic outcomes has shown to be a very productive way to overcome the limitations of scarce long–term longitudinal data.[4] We study social groups from different historical periods and ethnic backgrounds linked with a clear position on the social ladder in the past, to evaluate their position in the educational system today.[5] Our analysis is based

[3]The *encomienda* was a formal institution established in different Latin American countries during the Spanish colonial period. It was the right of Spaniards living in the colony to collect tribute from the Indigenous communities.

[4]A whole branch of literature about social mobility and persistence has been built around the use of surnames. For an extensive review of name-based methods see Santavirta and Stuhler (2020). Some relevant examples of fruitful applications of these methods in different historical and geographical contexts are summarised in Clark (2014) and Clark et al. (2015).

[5]Our historical groups are: Indigenous pre–Hispanic groups, 17$^{\text{th}}$ century Spaniard colonial officers (*encomenderos*), African enslaved people, members of privileged families with access to higher

on the use of rare surnames to follow multiple generations and take advantage of the availability of rich contemporary administrative data–sets and historical information (Clark et al., 2015; Guell et al., 2013; Jaramillo-Echeverri, Álvarez & Bro, 2021). Also, we exploit the official rule of registering both paternal and maternal surnames, inherited from the Spanish naming customs to describe the marriage patterns of our historical groups in the educational system. Finally, following the recent work by Chiappori, Dias and Meghir (2020), we measure assortative mating in the marriage market as another measure of segregation in the educational system.

We find significant differences between historical groups in different educational categories (e.g. low-quality education or high-prestige high schools). These differences confirm the exclusion of the ethnic groups and a systematic pattern of segregation beginning with access to exclusive private schools for the elites, from which Afro-Colombians and Indigenous are excluded. These differences also reflect different patterns of persistence. The encomenderos of the 17^{th} century and the slaves owners of the 19^{th} century seem to have lost their high-status position as these groups do not appear over-represented in high-quality education. On the contrary, the elites related to the emerging bourgeoisie of the end of the 19^{th} century, present strong persistence of their privileged high status. The social status of those families who accessed high quality and high prestige education during the 17^{th} to the early 19^{th} centuries show weaker persistence in comparison to the post-1870s elite, but they still appear overrepresented in elite-type institutions. Finally, historically excluded social groups (Afro-Colombians and Indigenous) show the lowest social status in all the educational categories we use to measure it. We conclude that the educational system in Colombia reproduces patterns of social exclusion that are rooted in the past and that it partially reflects the persistence of colonial segregated social structures.

education in the 17^{th} and late 18^{th} century, slaves owners in 1851, and members of different social and business elites of the late 19^{th} and the beginning of the 20^{th} centuries.

When we observe the marriage market in the contemporary educational system the results show there is virtually no matching between the historical groups. This is especially noticeable in elite institutions like Los Andes University, while intergroup marriages are less rare in low-quality public schools. These results confirm the strong link between assortative mating and social mobility. If schools are one of the most significant spaces of social interaction, then early low or nil exposure to social interaction between social groups can contribute to perpetuating barriers to social mobility.

Many contributions point to the persistence of colonial institutions in Latin America and their effects on contemporary economic performance (Sokoloff & Engerman, 2000; Acemoglu, Johnson & Robinson, 2005; Williamson, 2015; Dell, 2010; Valencia Caicedo, 2019). For Colombia, in particular, there are several analyses at the sub-national level, showing the importance of certain forms of colonial institutions as determinants of long-term differences in development in different regions of the country (García-Jimeno, 2005; Acemoglu, Bautista, Querubin & Robinson, 2007; Acemoglu et al., 2012; Maloney & Valencia Caicedo, 2016; Kalmanovitz, 2001; Alvarez, 2018). Most of these studies conclude that the presence of extractive institutions that were based on forced labour was detrimental to the long-term development of the regions. For example, Acemoglu et al. (2012) find that the presence of gold mines in the 17^{th} and 18^{th} centuries is associated with contemporary lower school enrolment and higher land inequality at the municipal level. We contribute to this literature by showing that the long-term effects of the colonial institutions were uneven across population groups. We find that Indigenous and Afro-Colombians which were at the bottom of the social ladder during the colonial period appear at the bottom in the contemporary educational system, while the elite status of the encomenderos and slaves owners, on the contrary, has not persisted over time.

Our research also sheds light on the origins of social segregation in the educational system. The results suggest that the elite of the early republic has been the most successful in maintaining its social status over time and indicate a substantial break with the colonial past as encomenderos and slaves owners seem to have lost their elite status. However, the findings also uncover a group of families that have been able to maintain their high status from colonial times to the present, as these surnames appear in several elite groups. This finding suggests a reinvention of the elite through history.

On the other hand, the bottom part has continued trapped with very scarce possibilities to move upward. This is also evident in the homogamy patterns we find. In line with previous research, we show that assortative marriage reflects the low social mobility of the country, particularly of those at the bottom of the social ladder. We contribute to the literature by providing a new picture of the marriage market in the country.

Additionally, we contribute to the literature on the case study of Colombia, enlarging the perspective on the long–term evolution of relative social status rather than focusing on contemporary short–term intergenerational correlations. The findings contribute to the literature by showing that educational systems matter in the perpetuation of the elites (Cardenas et al., 2021; Clark & Cummins, 2013). Finally, we contribute to the literature by assembling several data–sets to track persistence and social mobility in the presence of fragmented information. In the future, we aim to compare Latin American countries systematically using surname-based methods to measure long-run social mobility.

5.2 Context

In this section we present a brief historical context since the arrival of the Spaniards to what we call the beginning of the modern era in the early 20^{th} century. We also discuss how the contemporary educational system is divided between public and private schools and what are the implications of this division in terms of opportunities to access high-quality education.

5.2.1 Historical legacy

After the first Spaniards settled in Colombia around 1500, the Colonial power established the institution of the *encomienda* in which the notable Spaniards living in the colony were allowed to collect tribute such as gold, fabrics, emeralds, and labour from the Indigenous communities in exchange for Christian education and protection (Colmenares, 2007). Soon after, the Indigenous were also forced to work in gold mines, *haciendas* and craft shops. By 1560 the *encomienda* was well-established in all the territory of Nueva Granada, but several social disagreements between the Crown and the encomenderos, and the demographic collapse of the Indigenous population caused its abolition at the end of the 17^{th} century (Kalmanovitz, 2008). Despite the extractive aspect of this institution, several authors have suggested that the *encomienda* had persistent positive effects on the long-term development of the municipalities in which this institution was present and that its presence is associated with lower levels of poverty and higher secondary school enrolments although it appears to have no effect on inequality (García-Jimeno, 2005; Faguet, Matajira & Sánchez, 2017; Mora Bustamante, 2016). This literature suggests that the causal channel of this persistence runs through the strengthened local presence of a more capable state since the *encomienda* was the precursor of the state. However, others have also pointed to the persistence of pre-colonial prosperity that is reflected in contemporary higher

incomes (Maloney & Valencia Caicedo, 2016).[6]

The striking demographic collapse of the Indigenous population during the 16th and 17th centuries resulted in the importation of enslaved people from Africa to work in the gold mines. By the 18th century the economy of Colombia was based on mining, agriculture and commerce, and in turn, these activities were established around slave labour. The regions historically associated with gold mining and the production of sugarcane and other forms of plantations were the preeminent destination of African enslaved people, mainly in the departments of Cauca, Antioquia, Chocó and Bolivar while most of the African enslaved people entered through the port of Cartagena (Jaramillo Uribe, 1963). As with the *encomienda*, researchers have measured the long-term impact of slavery on economic development (García-Jimeno, 2005; Acemoglu et al., 2012) concluding that the presence of gold mines in the 17th and 18th centuries is associated with contemporary lower school enrolment and higher land inequality at the municipal level. Slavery was finally abolished in 1851 and the Afro-descendant populations founded several communities in the vicinity of the Caribbean and Pacific coasts.

Beyond slavery, one of the salient traits of the segregation patterns during the colonial period was the use of blood purity. This royal legislation, that originated in Spain during the 15th century to exclude Jews and Arabs from important public positions, was established in the colonies as a mechanism to avoid racial mixing and to maintain control of the region's political and religious institutions (Villalobos, 2005). The legislation required proof of the purity of blood, legitimate birth and honorability for appointment to several positions, including ecclesiastic and military, and also admission to secondary and higher education (Helg, 2014). Schools and higher-education institutions appeared during the second half of the 16th century to

[6]The *encomienda* settled in more dense areas where the presence of the pre-colonial population was higher.

educate the heirs of Spaniard families. These institutions were the main educational centre of the ruling elites during colonial times but also of the leaders of the bourgeoisie pro-republican groups during the Independence process that started after 1810.

Independence movements against the Spanish Crown took place in the first decade of the 19th century, and the consolidation of an independent republic was achieved in the 1820s. The causes and effects of the Independence wars in Latin American countries during the first half of the 19th century are still a matter of open debate in the literature (Grafe & Irigoin, 2006). For example, it is argued that the fiscal system was destroyed while internal conflicts (wars) increased in the regions (Coatsworth, 1993; Bulmer-Thomas, 2003). A good part of the Colombian historiography coincides with pointing out that the break with the Spanish monarchy did not constitute a radical change in the political, economic, and cultural elites of these countries, as regional elites defended their political and economic interests. Ripoll Echeverría (2006) calls it a "revolution without renovation", while Olarte (1993) points out that the Creole elites (whites from Spanish families born in the colonial territory), who led the Independence movements once Napoleon overthrew the King of Spain in 1808, consolidated a highly hierarchical system, maintaining the social class segregation of the old regime.

In the long-run, Latin American countries also benefited from institutional and economic modernisation (Coatsworth, 1993; Prados de la Escosura, 2009). The importance of the slave economy started decreasing as new industries emerged, while mining lost importance in comparison with new agricultural products (Bértola & Ocampo, 2012). The expansion of agricultural exports was accompanied by rapid colonisation of new lands in the lower slopes of the mountain ranges and in the interior of the country that continued until the beginning of the 20th century. The expansion of the tropical export crops and especially the production of coffee laid the

foundation of the modern economic era with the emergence of early forms of industrial development in Antioquia and Cundinamarca and the banking sector in 1870. However, the persistence and the economic effects of the new institutions and elites that emerged after the independence have been less studied in the literature. But by the beginning of the 20th century, an economic and cultural elite had managed to consolidate in some cities such as Bogota, Medellin, and Cartagena. These elite groups were linked to productive activities such as the financial and commercial system and also promoted the creation of cultural (but exclusive) spaces such as social clubs for gentlemen. As we will discuss in the next section, these groups also established high-quality *private* schools and universities that contributed to the persistence of a highly segregated society.

5.2.2 Education in Colombia: The present

Today, the educational system in Colombia displays symptoms of social segregation as shown by Cardenas et al. (2021). This phenomenon has been also observed in other countries like the US. In fact, Chetty, Hendren, Kline and Saez (2014) and Chetty, Friedman, Saez, Turner and Yagan (2017) present compelling evidence of inequality of opportunity in access to high-quality education and its effects on the US schooling system. Both in the US and in Colombia these studies conclude that inequality in access to high-quality education limits the ability of education to promote intergenerational mobility.

Since the last constitutional reform (1991), access to primary and secondary schooling in Colombia has enlarged. Coverage rates in primary education are above 90% and in secondary education are close to 75%. Also access to higher education has improved during the last decade, going from around 35% (2010) to 52% (in 2020). However, the quality of education is still poor and unequal (Camacho, Messina & Uribe Barrera, 2017). In fact, Colombia has the lowest scores in the PISA stand-

ardised tests among the members of the OECD, and it is below the average level in South America.

Beyond improvements in coverage, the most salient characteristic of the educational system is the heterogeneity in opportunities to access high-quality institutions. In a country with an average low level of quality, the existence of a small number of high-quality secondary schools, located mainly in the urban areas, and especially in Bogota, shows evidence of segregation and status-driven access to those schools. This generates a bottleneck in access to higher education with a clear bias that favours access by higher-income households, and generates social apartheid as Cardenas et al. (2021) called it.[7] The probability to access higher education for a member of the lower 25% of the income distribution is around 38% while it is near 45% for a member of a household in the top 25% (See Cardenas et al. (2021) page 99). Furthermore, García Jaramillo, Rodríguez Orgales, Sánchez Torres and Bedoya Ospina (2015) estimate that the probability of accessing higher education depends strongly on the level of education attained by the mother. The intergenerational correlation between mothers and children who access higher education is 92.4%, while this same correlation is 19.6% for mothers who only complete primary education and young people who only complete primary education.

Additionally, school quality is strongly linked with the public/private character of the institution. This implies that social status is correlated perfectly with school performance. However, educational quality is not the only feature offered by private schools. There is a clear elitist bias and a social status signalling mechanism related to attending a private instead of a public school.

Public schools tend to perform worse on average than private schools on different measures of quality. But there are a significant number of private schools

[7] A similar pattern is observed in the US where college access varies greatly by parent income. See Chetty et al. (2017).

that have lower quality performance than public schools with exceptionally good performance. In the midst of this heterogeneity, Cardenas et al. (2021) show that there is a marked preference among the middle classes (especially the lower-middle income classes) to make an extraordinary economic effort for their children to attend private schools, even those of similar or lower quality than public schools where their children could attend for free. The work of Fergusson and Flórez (2021b) shows in detail how there is evidence of an elitist bias in school preferences, which explains these decisions. More precisely, private schools tend to seek to imitate, both in their names and in their educational proposals, the characteristics of very high-quality private schools. These high-quality schools are mostly international schools or schools that offer bilingual education. The top of these schools is located in the capitals of the richest regions of the country.

The causes behind this educational segregation are unknown. Cardenas et al. (2021) suggest that at least since the early-republic, political and ideological tensions prevented the advancement of public education. The limited public education created differences in social capital that can explain today's educational apartheid. In this chapter, we test if these historical roots go also back to colonial times when colonial institutions required blood purity to enter higher education.

5.3 Data

5.3.1 Surnames

Measuring the persistence of social status in the long run in Colombia is difficult because of the lack of continuous and reliable historical administrative sources. For the 20^{th} century, censuses available to researchers do not include names, which makes it difficult to follow individuals over time. Censuses of the colonial and early republic periods are dispersed in several archives and have been not digitised. Longitudinal

surveys are scarce in the country, and most of them are very recent with few waves. As an alternative method to follow the evolution of long–term social status in a multigenerational framework, the literature proposes the use of rare family surnames to measure persistence and social mobility.[8] The use of surnames is relevant as surnames are almost always passed from one generation to another carrying relevant characteristics that are strongly related to social status.[9] The assumption behind the rarity of the surnames is that sufficiently rare surnames in a population represent observations of the same familial dynasty.

Since, at least, the Bourbon reforms (18^{th} century), the modern use of the Spanish naming system was adopted in Colombia (*Viceroyalty of New Granada*) and during the 19^{th} century acquired legal enforcement. This system consists of the customary or legal use of two surnames after a given name. The first surname is the father's first surname and the second surname is the mother's first surname. This implies that both family lines are present in a given name.

In Colombia, this naming system was conserved and reinforced after independence from the Spanish Crown. For our purpose, this system contributes to the preservation of the record of the mother's surnames for, at least, one additional generation compared with naming systems conserving only one surname (usually the father's). Instead of losing the mother's surname, we can identify both family links by observing contemporary cross-sectional data.[10]

To define a set of sufficiently rare surnames in contemporary Colombia, we use a database that contains all the students enrolled in third grade of primary

[8]See for example Guell et al. (2007) and Clark (2014). See also Barone and Mocetti (2021) for an application of surnames in the historical context of Florence.

[9]For example, Guell et al. (2013) show that rare surnames correlated with observable metrics of social and economic outcomes for the Spanish case.

[10]Usually the first surname of all children is the paternal surname followed by the maternal surname. In the 1990s it was introduced the possibility to adopt as a first surname the mother's surname. However, if both parents are known and officially registered as such, both surnames have to be included as part of the legal name of a person.

school in Colombia in 2016, corresponding to more than 90% of the families with children at this age (8-9 years old).[11] This sample has the advantage of including a good representation of low, high and middle-income families. Using the Integrated System of Enrolment (SIMAT- Sistema Integrado de Matricula) we have access to the individual data for 997,036 students, including their full names, the schools they attend and the regions where they are enrolled. The frequency of appearance of every surname in the SIMAT measures the rarity and commonality and the distribution of these frequencies can be seen in Fig. 5.1.[12]

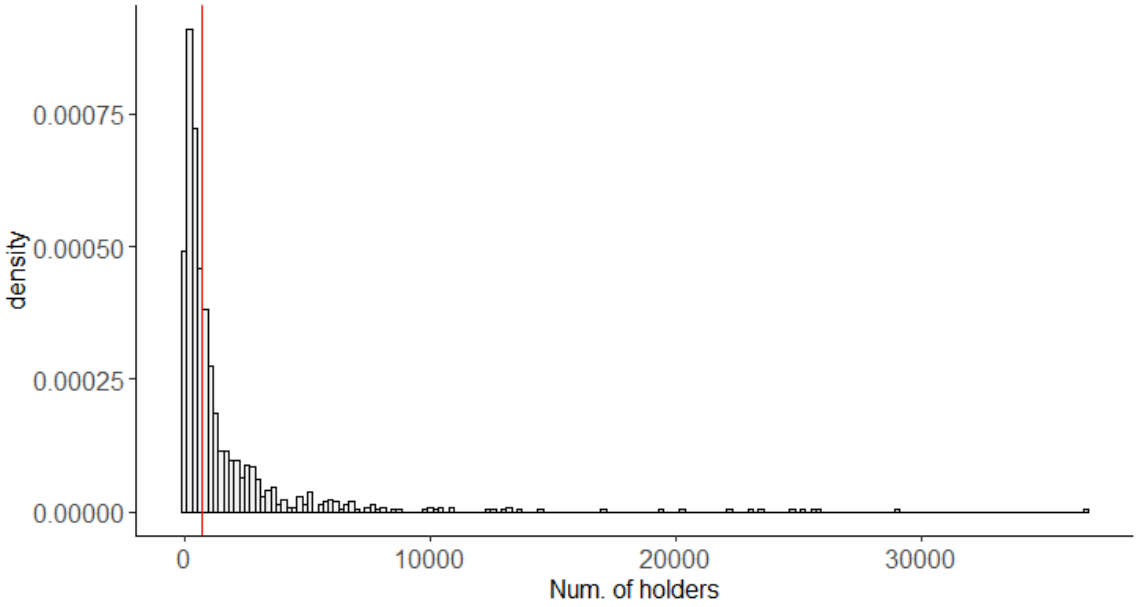
From this sample of Colombian surnames, we select those that have less than 2% of the number of holders of the most common surname in the sample (Rodriguez), and more than 3 holders.[13] The latter limit is set to avoid possible remaining misspellings registered as a one–holder's surname. This sample defines the list that we consider as being part of the same extended family, and therefore, belonging to dynasties whose status may have been persistent in the past.

[11]For primary school (1 to 5 grade) in 2016 the total net enrolment ratio in Colombia was 93,58 and the gross enrolment was 113.56, according to the National Ministry of Education.

[12]The pattern of this distribution of surnames is well established in the literature (Santavirta & Stuhler, 2020) and it is worth noticing that this common feature of the distribution of surnames across *Western* societies presents an advantage for the method because the number of rare surnames is large.

[13]We modify the definition of rarity using different thresholds and we do not find significant differences in our results. See results in the appendix

Figure 5.1: Distribution of surnames in Colombia, 2016



Note: Figure 5.1 shows the distribution of paternal and maternal surnames. We find 9,117 surnames held by only one student, such as *Savatoni* or *Abacuk*. The last point in the distribution is the surname *Rodriguez*, the most common surname in Colombia, which appears 36,678 times in the dataset. The red line indicates the threshold of our preferred definition of rarity. Source: SIMAT, 8-9 years old enrolled in third grade of primary school, 2016.

We define rarity as a proportion of the most common surname as a way to define a standard of reference to navigate across different data sources. If our dataset of reference (SIMAT) is a good representation of the distribution of surnames in the whole population, keeping a reference to the frequency of Rodriguez holders, as a measure of rarity or commonality of the surnames seems to be a reasonable strategy.[14] The total number of rare surnames obtained is 9,902, and we also include in our data–base the 20 most common surnames as a comparison group.

As reported in Table 5.1, from a total of 997,036 students we obtained 23,144 unique paternal and maternal surnames. Some particular surnames needed a slight transformation to avoid confounders or typos. We transformed these surnames by

[14]It is worth noticing that, according to the national official registry (*Registraduría Nacional del Estado Civil*), holders of Rodriguez represent around 1.5% of the entire population in Colombia. In our sample, this surname represents 1.8% of the total.

removing accents or special characters (ñ, ü) and for the compound surnames like "De la Cruz", we merged them as a unique word.

Table 5.1: Surnames summary statistics

Number of distinct names	23,144
Mean frequency per name	86.16
% of people holding top-20 surnames	20%
% of people holding top-50 surnames	32%
% of people holding rare surnames	24%

Note: Rare surnames are defined as those held by more than 3 holders to avoid misspellings and less than 2% of the most common surname in the sample (Rodriguez). This is our preferred definition of rarity but we modify this threshold as part of the robustness checks.

5.3.2 Historical membership: elites and ethnic groups.

The next step towards the connection of past and present is to identify groups of surnames providing information about their historical social status. We are interested in those groups that occupy an extreme place at some end of the continuum of social status, at different points in history. We select two ethnic groups, four groups that represent the pre-industrial elite, and two groups that belong to the modern elite. For these groups, we have information on their historical social status linked to their surnames and we can follow them in the educational system today. We define a *group* as a set of families that share the same socioeconomic status measured, for example, by relevant ethnic social markers, access to elite education in the 17^{th}-18^{th} centuries, or owing African enslaved people at the end of the slavery in Colombia in 1851. We collect the surnames of these families using several sources, including primary and secondary sources. Table 5.2 summarises these groups and we describe in detail the historical groups, their economic and social position in history and the sources from which we collected the data.

Table 5.2: Historical groups

Social condition	Historical group	Period	Characteristic
Ethnic groups	Indigenous	Forced work since 15^{th} century	racial/ethnic group
	Afrocolombian	Forced work since 16^{th} century	racial/ethnic group
Pre-industrial elite	Encomenderos	16^{th} – 17^{th} century	Colonial elite
	Slave owners	Received compensation from the state in 1851	Colonial elite
	Colegio San Bartolome	1605 – 1820	Educational elite
	Colegio del Rosario	1773 – 1842	Educational elite
Modern Elite	Bank founders	1870 – 1885	Business elite
	Jockey Club	1874 – 1902	Social elite

Note: Table 5.8 presents the number of surnames in each list of historical groups and the number of surnames we define as rare using the number of holders in SIMAT.

The Colonial legacy: 16^{th} - 19^{th} centuries

Encomenderos

The *encomienda* was a formal institution that Spain established in different Latin American countries. It was the right of notable Spaniards living in the colony to collect tribute such as gold, fabrics, emeralds, and labour from the Indigenous communities while offering them religious education. In Colombia, the size of one *encomienda* ranged from 100 to 3,900 Indigenous per *encomienda*, and in the regions where there was a higher presence of agricultural societies, there was a higher number of *encomiendas*, creating significant variation in the wealth of an encomendero. The *encomienda* lasted from 1550 to the end of the 17^{th} century when the demographic collapse of the Indigenous population made it unsustainable (Kalmanovitz, 2008). The encomenderos were a heterogeneous group and their social status changed significantly from 1560 to the moment when the *encomienda* was abolished. But these first "recognised" settlers in the colony had access to the best resource available in the colonies and they created the first economic elite of the 17^{th} century. After the collapse of the *encomienda*, forced labour continued to be part of the economic elite of the territory, but other activities like commerce and trade emerged in the colonies, and a political elite was established by the 1750s.

Our first list of surnames comes from an official document written in 1674 by Juan Florez Ocariz. It is an extensive report about the people living in that period in the New Kingdom of Granada. This text, entitled *Genealogia del Nuevo Reyno de Granada*, contains information about various personalities that were related to the conquest and founding of the country. The book contains detailed lists of members of the military campaigns of occupation of contemporary Colombia and Venezuela, including also some parts of Ecuador and Peru. From Carrizosa (1990) we obtain the names and surnames of the people that appear in the original book with their original spelling.[15] Particularly, we focus on those that appear under the title of *encomendero*.

Indigenous

It is well-known that Indigenous populations in Latin America have been historically marginalised. As discussed above, soon after the arrival and establishment of the Spaniards the *encomienda* was established as a way of assigning Indigenous labour and goods to the Spaniards. For more than one century Indigenous communities had the obligation of sending workers to the mines (mita), the haciendas or the craft shops. After 1590 a series of changes took place in order to guarantee the safety of the communities. Among those changes, the Crown established the prohibition of using Indigenous for personal services or to provide food and crops for the encomenderos. As the reduction of the Indigenous population continued, the Crown decided to import African enslaved people to replace Indigenous labour, particularly in the mines, while organising the Indigenous population in small shelters near the cultivated lands. This division in the territory laid the foundations of a remarkably hierarchical society in which Indigenous groups kept losing part of their original territories. To reverse this process between 1988 and 1990, approximately 20% of the national territory was given to the control of communities that represented less than 2% of the population

[15]This means that some of the surnames like Ximenez that have been transformed in time to become Jimenez appear in our list of genealogies, but we cannot link them with their contemporary spelling.

(Melo, 2017).

In Colombia surnames such as Cárdenas, Tobar o Cuero which are Spanish in their origin, are common in some Indigenous communities. However, other communities have retained their pre-hispanic surnames as Sastoque or Botina. The official use of surnames among Indigenous communities appeared during the 16th century thanks to the adoption of the institution of *resguardos* in the Viceroyalty of New Granada. Similar to the congregaciones in New Spain, the *resguardos* allowed for a certain autonomy on the exploitation of land for an Indigenous family or group of families outside the territory of an *encomienda*. The main goal of this institution was to maintain control over Indigenous groups. They were forced to be baptised and to adopt a Christian name to which the original given name of the heads of the family, usually men, was added as a surname. On other occasions, surnames were assigned to the whole group based on a topographic characteristic of the region. We can use surnames to track families over time because in most Indigenous groups the patrilineal inheritance in which the children inherit the surnames from the parents is preserved (Instituto Colombiano de Cultura Hispanica, 1992). In this research, we use Indigenous names only, since in this way we can link the surname with its historical social status. We have collected a list of 131 Indigenous surnames from a detailed anthropological study (Instituto Colombiano de Cultura Hispanica, 1992).

Slaves owners

Between the 16th and the 17th century, the Crown started to import African enslaved people to substitute for the declining Indigenous population. By the 18th century the economy of Nueva Granada was based on mining, agriculture and commerce, and in turn, these activities were established around slave labour. However, after independence, the importance of the slave economy started decreasing as new industries emerged, and mining lost importance in comparison with new agricultural

products. Likewise, international pressure made the manumission process irreversible. In 1851 Congress issued the law of May 21st, 1851, by which the African enslaved people would be free from January 1st, 1852, and the slaves owners would be compensated with bonds of the state.

Tovar (2007) and Tovar and Tovar (2009) compile the list of the slaves owners who were covered by this law and received compensation from the state. We use this list of 1,667 surnames of slaves owners that received this compensation between 1851 and 1852. This list represents the elite of the Colonial legacy that by 1851 was in decline. Several caveats should be made with this list, and with the social status that these surnames can capture. This list does not give us a consistent social status, as we do not have complete information about the number of African enslaved people for each individual. Tovar (2007) present a list of 30 people who had more than 50 African enslaved people for those who he could assign the number of African enslaved people. For example, in this list, Napoleon Lozano appears with 196 African enslaved people, while Ramon Cabezas with 52. The variance in the number of African enslaved people may indicate variance in social status, as well as variance in the capital received by the state. It is not clear then if this elite continued to be an elite group after they freed the African enslaved people, or if on the contrary, this elite was already in decline. Similarly, an important part of the Colombian historiography suggests that slavery was only profitable in very exceptional cases (Kalmanovitz, 2008). An additional caveat regarding the social status captured by the surnames is that after the process of manumission many of the African enslaved people took the surname of their owner, and therefore surnames like Mosquera may be related to the black population, a population that has been systematically marginalised in the country.

Afro-Colombians

The Afro-descendant population in Colombia has its origins, in its immense majority, in the African enslaved people that arrived during the 16th to the beginning of the 19th centuries. The regions historically associated with gold mining and the production of sugarcane and other forms of plantations were the preeminent destination of black African enslaved people, mainly in the departments of Cauca, Antioquia, Chocó and Bolivar (Jaramillo Uribe, 1963). During the colony, the African enslaved people were at the bottom of the social ladder, and flights and rebellions were common. After the Independence process and the definitive abolition of slavery in 1851, Afro-descendant populations founded several communities in the vicinity of the Caribbean and Pacific coasts. A considerable part of the African enslaved people adopted the surnames of their owners at the moment of gaining freedom, but the population of these regions sought to reconstruct their African ancestral identities, or at least avoid identification with the former slaves owners. These populations thus adopted surnames that corresponded to topographical surnames, ethnic names or designated physical characteristics of their ancestors before being Christianised or officially being given a Spaniard or Mestizo surname.

Therefore, in Colombia, a set of surnames that are properly Afro-descendant and that have almost zero frequency of occurrence among non-Afro-Colombian populations remain to this day. Using the main anthropological studies that have sought to define inventories of these typically Afro-descendant surnames, we compiled a list of 276 of them. We use two lists of surnames compiled by afrocolombianists, built upon a set of past studies (Zapata Olivella & Mina Aragon, 2014; Mosquera, 2014). They discuss and group different ethnographic scholarly works, in particular the seminal contributions on this subject by Velásquez-Murillo (1962) and Arrazola (1970). We then compared these lists of surnames with our other groups to avoid coincidences with surnames of slave–owners (as the very frequent case of the surname Mosquera).

Elite schools and universities: 17th century

During the Spanish colonial period (mid-16th to mid-19th centuries) and until the republican reforms of the mid-19th century, the access to high school and higher education was ruled by a selection process based on noble heritage and "purity of blood". Only the heirs of Spanish families, and in particular those with noble privileges were admitted to the "colegios mayores" (higher education). The selection process was based on the examination of candidates' files, and recommendation letters from prestigious acquaintances of the candidate. An important part of those archival materials was conserved. Two of the most important, and most ancient, institutions of the Colonial era exist today: *Colegio Mayor de San Bartolomé* and *Colegio Mayor de Nuestra Señora del Rosario*.

Colegio San Bartolomé

The *Colegio San Bartolome* is a private primary and secondary school founded in Bogota in 1604, and since these times it has been a flagship in high-quality education. It was founded by archbishop Bartolome Lobo Guerrero to educate Spaniards, born out of a legitimate marriage, who already knew how to read and write. In particular, the descendants from Conquistadores were preferred (Mejía, 1996). It is the oldest school in Colombia and one of the most prestigious institutions from which 28 presidents of Colombia have graduated.

We collected a list of 224 students of the Colegio San Bartolome during the colonial period until the Independence, using the book *Real Colegio Mayor y Seminario de San Bartolome, 1605 to 1820* published by the Instituto Colombiano de Cultura Hispanica in 1996. This book accounts for the history of the school from its foundation until the Colombian Independence.

Universidad del Rosario

The *Colegio Mayor de Nuestra Señora del Rosario*, today is known simply as Universidad del Rosario, is a higher education institution founded in Bogota in 1653. It is one of the oldest universities in Latin America and the oldest with uninterrupted functioning in Colombia since Colonial times. From the 17th to the 18th century Universidad del Rosario was the most prestigious higher education institution in the country. It was ruled under the Dominican Order and proof of blood purity or nobility (*hidalguia*) was necessary for gaining access to it. As a result, this institution was the main educational centre of the ruling elites during colonial times. Philosophers, lawyers, theologians and physicians were formed at this university.

We established a list of 5,482 male students of the Universidad del Rosario using a secondary source: a two volumes work by Maria Clara Guillen de Iriarte (Guillén de Iriarte, 2006, 2008) who reports a list of all the graduates from 1773 until 1842. This work contains information about the students, their parents and eventually their patrons. The books offer a reliable transcription of the full set of manuscript archival sources of curricula of graduates.

The beginning of modernity: late 19th and early 20th century

Jockey Club

By the end of the 19th century, aiming at emulating the British elite's practices, the Colombian elite, in particular in Bogota, founded social and sports clubs. One of the most exclusive social clubs is, still today, the *Jockey Club*.[16] The purpose of founding the Jockey Club was to create a private space for social gatherings for its members. When the Jockey Club started, "the members of the Club were businessmen, renters and heirs, partially dedicated to the agriculture of grains (wheat and

[16]This social club and its members, had such influence on the country's politics and economic development during the 19th and 20th centuries that some called it "the other Casa de Nariño", meaning the other presidential house. This wordplay comes from the fact that the President's residence is named after one main figure of the independence process: Antonio Nariño and the original building where the Jockey Club was established was Nariño's house.

barley), coffee growers, milkmen, promoters or entrepreneurs of public services, senior politicians, some humanists, literati, lawyers, doctors and a few other professionals" (Davila-Ortiz, 2010). This club has been one of the most exclusive institutions in Colombia since its foundation. Currently, the Jockey Club continues to be a flagship institution of the Colombian elite. There are no memberships available, and those who want to be part of the group must wait for a member to retire or die without leaving an heir to take their place, which is indicative of the exclusiveness of the club. Even more, the number of members who can make decisions and be elected to management positions is limited to only 600 people.

We collected a list of 811 individuals who were associated with or were founders of this elitist (and all-male) institution. From the founding acts of the Jockey Club, we obtained a list of the 123 men that participated in its foundation on September 20th, 1874. Also, we have the list of the associates in 1875, and the following "re-foundations" acts of 1888, 1890 and 1902.

Bank founders

A period of free banking took place in Colombia between 1870-1885. During this period private banks could issue banknotes, backed by their reserves of silver and gold. In total 36 banks were established in different regions of the country. These banks were opened mainly by merchants from some regions in search of more liquidity. They also intended to link the banking system with local and international trade. Likewise, "the creation of banks in the most prosperous regions of the country allowed the regional and local elites to control the monetary and credit system, consolidating their power over the financial market" (Correa, 2009).

We have a list of the founders of the first 12 banks established in the country, between 1870 and 1883, from the foundation letters of each firm. This list includes the 3 biggest banks of the period: *de Colombia*, *de Bogota* and *de Antoquia*. The former

two are still among the biggest private banks today. So we can observe a business elite of the period, as these firms were the first forms of stock-holding corporations in the country and signal the first steps of economic modernity well before industrial development (towards the 1920s).

5.3.3 Relative Representation in the educational system

After constructing our sample of rare surnames, we focus on the opportunity to access education in contemporary Colombia. We use different administrative datasets collected at the individual level that contain both surnames and relevant information on the quality of education, depending on the school or university the individual attended. To define contemporary access to education we construct eight educational categories (e.g. high and low quality education, public low quality education and high prestige high schools) that are described in Table 5.3. Appendix Section 5.A.1 describes each dataset in detail, the sources from which we obtained them and provides additional information about the educational system.

Table 5.3: Description of contemporary social category – Access to education

Social condition	Outcome	Definition
High status	Los Andes University	Most prestigious private university in Bogota
	International Schools	Most prestigious and expensive schools located in Bogota
	High Quality school	school being on top 5% of the score in the national standardised test
	HQ Public school	Public school being on top 5% of the score in the national standardised test
	HQ Private school	Private school being on top 5% of the score in the national standardised test
Low status	Low Quality school	school being on bottom 5% of the score in the national standardised test
	LQ Public school	Public school being on bottom 5% of the score in the national standardised test
	LQ Private school	Private school being on bottom 5% of the score in the national standardised test

Note: The table summarises the variables we define as social categories, and presents a description of their construction. The variables come from the registration forms filled out by students who finish high school and take a standardised test, known as SABER 11, similar to the SAT in the USA or the A-levels in Europe. The variable of Los Andes comes from the graduation act of the University. See Appendix Section 5.A.1 for more details.

With this information, the relative representation of any surname j in each period is calculated as described in equation 5.1.[17] Then, we go from individual-

[17]It is important to note here that equation 5.1 refers to the relative representation of an individual surname. This can be used for any surname found in the contemporary socioeconomic outcome in the corresponding dataset. With the RR_j we can estimate the average level of the relative representation

level data to surname-level data by taking the average relative representation of each surname.

$$RR_j = \frac{\text{Share of j in access to education}}{\text{Share of j in SIMAT}} \tag{5.1}$$

This measure captures how much a surname is over-represented or under-represented in a specific outcome given its frequency in the total population. For the common surnames, which tend to behave similarly to the average population, the relative representation in any period and in any social outcome will not be statistically different from 1. For high-status surnames in high-status outcomes, the relative representation of the surname will exceed 1, but in low-status outcomes, it will fall below 1. On the contrary, for low-status surnames, the relative representation will exceed one in low-status outcomes, and be lower than 1 in high-status outcomes.[18] With this measure it is possible to assess how persistent is historical status.

5.4 Methodology

5.4.1 Persistence of historical groups

To evaluate whether the contemporary differences in the educational system have deep roots in the past we regress the relative representation at the surname-level (RR_j) on a set of dummy variables for each social group. We test if the mean level of RR of rare surnames belonging to historical groups is statistically different to the mean value of this metric for the most common surnames following equation 5.2.

of a historical group. The historical social groups are used as a marker of social status in the past and can be considered as an attribute of a surname.

[18] We calculate the relative representation at the surname-level for each educational outcome, and we find that its distribution is clustered around zero and it has a long right-hand tail as shown in the Appendix 5.11

$$Z[RR_j] = c + \sum_{g=1}^{n} \delta D_g + \varepsilon_j \tag{5.2}$$

Where g is the social group to which surname j belongs, D_g is the matrix of dummies for each group g and $Z[RR_j]$ is estimated at the surname level for each educational outcome, and c is the constant term. The standard errors are robust to account for potential heteroscedasticity given that it is likely that like other social status measures, the variance of the Relative Representation is different across the historical groups. To allow for an easier interpretation of the results, we standardise the RR_j values using a *Z-score* method and we will refer to this transformation as $Z[RR]$. This transformation allows us to interpret the coefficients of our regressions in terms of standard deviations with respect to the mean value. As previously discussed, a common surname is expected to be equally present in any social condition in a similar proportion as it is within the entire population and it has been largely documented that the average social status of common surnames, in Western cultures, is similar to the average social status of the society studied (Clark et al., 2015; Chetty et al., 2014). Given that the most common surnames have an average social status, the value of the standardised relative representation (RR) of those surnames (i.e. Rodriguez) is close to 0, or at least not statistically different from 0.

5.4.2 Assortative mating

The literature has identified assortative matching in the marriage market (also known as homogamy or simple assortative marriage) as a key mechanism in the persistence of social status because it reflects the degree to which individuals with similar socioeconomic characteristics marry each other (Blossfeld & Timm, 2003). The literature argues that homogamy allows elite social groups to maintain their social status for their future generations and can contribute to a segregating pattern reproducing the

exclusion of underclass groups and the evidence shows a strong correlation among spouses, by religion, ethnic group and social status, whether measured by occupation or education level (Blossfeld, 2009; Schwartz, 2013; Ermisch, Francesconi & Siedler, 2006). Nonetheless, empirical research discussing how rigid social structures can be maintained by assortative mating has been limited by the availability of data.

There are not many studies looking at assortative mating for Latin America, but one exception is the work by Dahan and Gaviria (2001) that studies the correlation of spouses' schooling for 16 countries, including Colombia, and its link with social mobility. Using household surveys of the late 1990s, the authors find that assortative mating levels are similar across countries and that assortative marriage and social mobility, and assortative marriage and inequality are strongly correlated. Not surprisingly, Colombia appears with the second-highest coefficient of assortative marriage (0.76) accompanied by one of the highest immobility indexes (0.59).

Our first contribution is to document marriage patterns by looking at the proportion of paternal-maternal surnames pairs belonging to each historical group. Exploiting the Spanish naming custom allows us to make use of both surnames (maternal and paternal) to observe marriage patterns as we consider each pair of surnames of any individual as a marriage.

First, we present a description of the marriage market in the whole population using the data from SIMAT and then we focus our analysis on three social conditions: low-quality public schooling, high-quality private schooling, and Los Andes University. The potential differences in the marriage market between these social conditions could be interpreted as segregated mating patterns. Then, to verify these patterns, we measure assortativeness by computing the Separable Extreme Value (SEV) index from Chiappori et al. (2020) which allows us to estimate a measurement of assortativeness that is comparable across different social conditions and by looking

at the index for each pair of historical groups we can determine if there are different levels of assortativeness between the groups.[19]

The index is based on an underlying structural marriage matching model and requires only 3 parameters: the proportion of elite paternal and maternal surnames (m and n respectively) and the proportion of marriages where both surnames are from an elite group (r).[20] If there was perfect assortative matching, then r would equal the minimum of m and n and the index will be higher than zero. If the matching was random, then r would equal the product of m and n, and the SEV index will be equal to zero. Table 5.4 exemplifies the different scenarios in the marriage market and Eq. (5.3) is the equation to calculate the index.

[19]Chiappori et al. (2020) highlight assortativeness as a local property, which means that it is possible for assortativeness to be high within one pair of social groups (e.g. Jockey members and Bank shareholders) but low within other pairs (e.g. Bank shareholders and Indigenous).

[20]To explain the construction of the index we will assume that we only have two categories: elite and underclass surnames, but in our analysis, we will use five categories: Indigenous, Afro-Colombians, pre-industrial elite, modern elite and commons.

Table 5.4: Assortative matching scenarios between groups

a. Observed matching

paternal/maternal	elite	underclass
elite	r	m-r
underclass	n-r	1+r-m-n

b. Random matching

paternal/maternal	elite	underclass
elite	mn	m(1-n)
underclass	n(1-m)	(1-m)(1-n)

c. Perfectly Assortative

paternal/maternal	elite	underclass
elite	n	0
underclass	0	1-n

Note: The Table 5.4 represents the different scenarios of matching with only two categories: elite and underclass surnames. Where r denotes the share of marriages where both surnames are from an elite group, and m and n denote the number of elite paternal and paternal surnames, respectively. In Panel C, we assume the surnames distributions for paternal and maternal surnames are identical. The table is based on Chiappori et al. (2020).

$$I_{SEV} = ln\left(\frac{r(1+r-m-n}{(n-r)(m-r)}\right) \tag{5.3}$$

To summarise, mn is the proportion of couples if randomly matched and there is positive Assortative Mating if $r \geq mn$. The I_{SEV} is the log of the ratio between the proportion of observed matches between couples of the same group and the observed proportion of couples formed by surnames belonging to different social groups. If this ratio is equal to one, then the index will be zero and we can conclude that there is random matching. If the ratio is higher than one is because there are more couples of the same group than couples formed by different social groups, which

indicates high assortative mating.

5.5 Results

5.5.1 Persistence of historical groups

Fig. 5.2 and Fig. 5.3 show the distribution of the average relative representation within each social group in contemporary educational categories. Fig. 5.2 shows this information for the 20 most common surnames in Colombia, and the figure can be used as the standard to compare our metrics of over or under-representation in a particular contemporary social status marker. A RR smaller than one implies that the group of surnames is underrepresented while a RR higher than 1 indicates that the group of surnames is over-represented in any given educational category. As shown in the graph, common surnames do not appear over or under-represented in any category, with an average RR close to 1.

Figure 5.2: Relative Representation in access to education - Common surnames

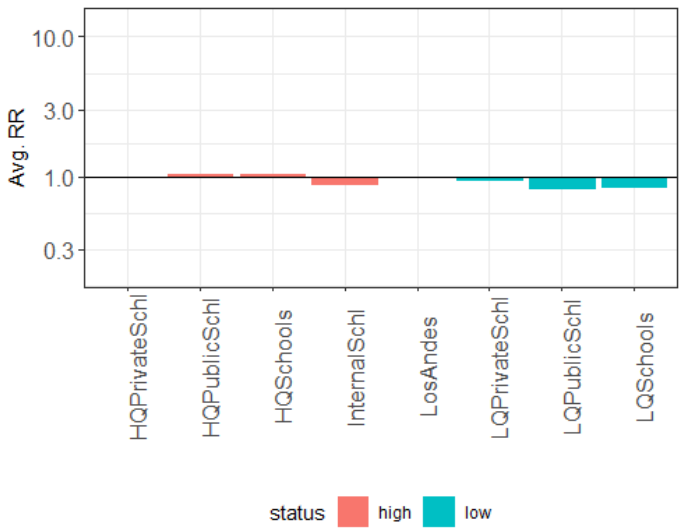


Fig. 5.2 presents the average relative representation of the 20 most common surnames in Colombia. The horizontal black line is set at $RR_j = 1$, corresponding to the neutral level of relative representation. The y-axis is log-scaled to ease interpretation.

The most striking result from Fig. 5.3 is the long persistence of the eth-

nic groups. The Indigenous and the Afro–Colombians remain underrepresented in high-status categories and this is especially pronounced in their exclusion from high-quality education and elite institutions. On the opposite side of the social ladder the Jockey Club, the banks–shareholders and, to a lesser degree, the graduates of the colonial schools, San Bartolome and Universidad del Rosario, show systematically over–representation in the high-status social categories while they are underrepresented in the low-status categories. Of special attention is the overrepresentation of the member of the Jockey Club in the graduates from Los Andes, as this group appears at a rate nearly 12 times higher than in the general population. Indigenous, on the opposite side, appear at an extremely low rate in Los Andes, about 0.3 times their frequency in the population. In comparison, Clark (2014) estimates that Ashkenazi Jews appear 6 times more among physicians in the US in 2012 than in the general population while Native Americans occur about 6% of the expected rate.

It is interesting to notice that the group of encomenderos seems to practically behave as the commons, as it does not appear systematically over or underrepresented in any category. These results reflect a process of convergence to the mean.[21]

Similarly, the group of slaves owners appears overrepresented in several high-quality outcomes as well as in all low-quality outcomes suggesting that the group lost its elite-type characteristic, indicating as well a regression to the mean process. This result is surprising given that Ager, Boustan and Eriksson (2021) found that in the case of the US, the sons and grandsons of former slaveholders surpassed their counterparts in educational and occupational attainment. However, it should be noticed that these two social groups are not only defined far back in time, but they

[21]In Jaramillo-Echeverri et al. (2021), we find a similar pattern for the Chilean case, where the Spaniard colonial officers appear as if they were slowly losing their elite status. We confirm that this group shows the lowest persistence of all elite groups in both Chile and Colombia when measured by the relative representation in high status categories. In Chile, they behave almost as the commons, only slightly over-represented in high income groups. In Colombia, this regression to the mean is more evident as they appear slightly under-represented in high quality schooling.

are also related to the pre-modern, and colonial society elites, while the other elite categories are not only recent but, as is the case of the graduates from the colonial schools, cover professional categories as doctors and leaders of the bourgeoisie pro-republican groups during the Independence process. In particular, we observe the slaves owners right at the collapse of slavery and it is likely that several of these families were unable to enrol in the new productive activities of the modern era.[22]

These results indicate that although status has persisted for certain groups, especially for Indigenous, Afro-Colombians and elite groups of the late 19$^{\text{th}}$ century, some pre-modern groups have lost their original high status. These findings suggest that the long-term effect of colonial institutions, such as the *encomienda* and slavery, were uneven across population groups.

[22]In Section 5.5.3 we show that these results are not driven by slaves taking slave-owners' surnames

Figure 5.3: Relative Representation in access to education

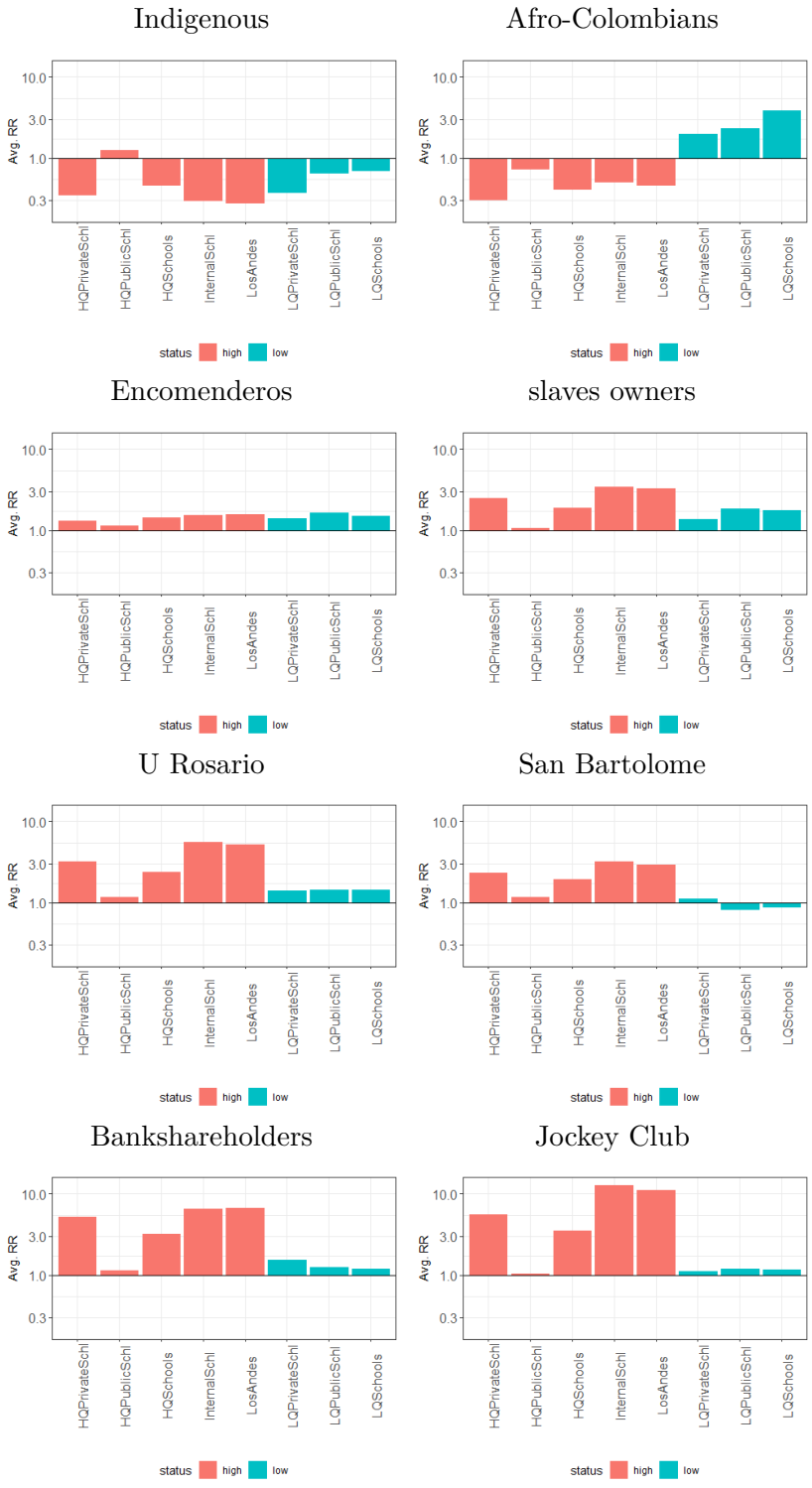


Fig. 5.3 presents the average relative representation of each group in the educational system. The horizontal black line is set at $RR_j = 1$, corresponding to the neutral level of relative representation. The y-axis is log-scaled to ease interpretation.

The results from table 5.5 confirm the patterns shown in figure 5.3. For example, looking at all High-Quality education measures, including Los Andes University, Indigenous and AfroColombian surnames appear underrepresented by around 0.5 standard deviations in comparison to the common surnames. On the contrary, members of the Jockey Club appear overrepresented by more than 0.7 standard deviations. In line with the findings of the previous section, Encomenderos are underrepresented by around 0.3 standard deviations while slaves owners and alumni from Colegio San Bartolome do not appear to be significantly different from the common surnames in these categories. Regarding the low-quality institutions, most elite groups do not behave significantly different from the commons whilst the Afro-Colombians are overrepresented by almost 0.6 standard deviations.

In general, the late 19^{th} century elite groups are over-represented in high-quality education, while the pre-industrial groups behave in all categories as common surnames, showing small and non-significant coefficients in almost all of the outcomes. On the contrary, Indigenous and Afro–Colombian surnames show a consistent pattern of exclusion from all of the privileged education outcomes. Surprisingly, in all low-status outcomes, such as low-quality public schools, the Indigenous surnames appear also underrepresented. One fact should be highlighted at this point, even though SIMAT covers around 90% of school-age children, according to the 2005 Census Report, the Indigenous population show the most lagging indicators in all the education variables, including enrolment in primary school. This implies that within the SIMAT the poorest Indigenous children are not even registered and therefore, we may be underestimating their participation in the lower strata. This could be a potential concern because this can bias the estimations upwards. However, as the results indicate high persistence, this implies that Indigenous could face even higher persistence.

For their part, the rare surnames that have not been identified in any historical group, are less represented in both ends of the educational outcomes in comparison with the commons. They appear underrepresented by around 0.1 to 0.3 standard deviations in comparison to the common surnames. This finding confirms that having a rare surname is not enough to predict a particular contemporary social status.

The results in Appendix Table 5.12 show that when we include all surnames the general findings hold, but, as expected the significance and explanatory power of the regressions fall considerably as the predictive power of frequent surnames is lower than the predictive power of rare surnames (Santavirta & Stuhler, 2020). On the contrary, when we impose a more restrictive definition of rarity (1 % of the most common surname) we find that the patterns remain but the significance of every coefficient improves as well as the overall fit of the regression, as shown in Appendix Table 5.13.

Table 5.5: Differences in educational categories by historical groups

	Los Andes	International schools	High Quality Schools	HQPrivateSchool	Low Quality Schools	LQPublicSchool
Ethnic groups						
Indigenous	−0.519***	−0.359**	−0.485***	−0.426***	−0.220***	−0.200**
	(0.145)	(0.118)	(0.079)	(0.092)	(0.070)	(0.067)
Afro-Colombians	−0.485***	−0.336**	−0.511***	−0.444***	0.568***	0.179
	(0.137)	(0.109)	(0.072)	(0.083)	(0.166)	(0.099)
Pre-industrial elites						
Encomendero	−0.390***	−0.298**	−0.191	−0.286***	−0.008	0.030
	(0.113)	(0.090)	(0.102)	(0.075)	(0.137)	(0.145)
slaves owners	-0.109	−0.069	−0.014	0.025	0.085	0.098
	(0.146)	(0.120)	(0.083)	(0.094)	(0.085)	(0.062)
URosario	0.212	0.171	0.137	0.132	−0.008	-0.011
	(0.178)	(0.154)	(0.119)	(0.125)	(0.07)	(0.064)
SanBartolome	−0.306*	−0.217*	−0.063	−0.127	−0.194***	-0.188***
	(0.127)	(0.110)	(0.080)	(0.091)	(0.054)	(0.061)
modern elites (late 19^{th} century)						
Bankshareholders	0.047***	−0.110	0.286	0.432	−0.067	-0.052
	(0.259)	(0.214)	(0.163)	(0.239)	(0.094)	(0.088)
JockeyClub	1.448***	1.320***	0.580***	0.756***	−0.063	-0.049
	(0.366)	(0.299)	(0.161)	(0.215)	(0.065)	(0.059)
Rare No Groups	−0.385***	−0.250***	−0.296***	−0.263***	−0.098	-0.078
	(0.144)	(0.117)	(0.075)	(0.090)	(0.064)	(0.058)
Constant	0.342*	0.219	0.260***	0.227*	0.085	0.068
	(0.144)	(0.117)	(0.074)	(0.089)	(0.063)	(0.057)
Observations	9,922	9,922	9,922	9,922	9,922	9,922
Adjusted R^2	0.044	0.029	0.020	0.024	0.005	0.002

Note: The equations are estimated using OLS.The 20 most commons surnames is the latent group of comparison.The RR_j is normalized. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

It should be noticed that it is possible that while Indigenous and Afro-Colomboans are exclusive categories, some elite surnames appear in more than one elite group. For example, the surname *Calvo* is the only surname that appears in all the elite groups, while the surnames *Cespedes, Plata, Valenzuela, Obregon, Quijano, Borda* and *Salgar* appear in 5 of the 6 elite groups. Of particular attention are those surnames that appear both in historical groups from the pre-modern era as well as from the late 19$^{\text{th}}$ century. These surnames represent the long-lasting elite that has persisted over time and that was able to reinvent itself, despite the failure of the *encomienda* or the independence process that sought to break the colonial institutions. To evaluate whether these long-surviving surnames have been more successful to maintain their social status in comparison to those that appear in only one period, we classify elite surnames into a new category *multiperiod elite* that includes all surnames that belong to at least one pre-modern group and one modern group.

Table 5.6 shows the results by educational categories. The coefficients of the *multiperiod elite* group show that in all high-quality categories this group appears around one standard deviation more represented than the common surnames and reflect the systematic overrepresentation of elite surnames in access to high-quality schooling. Similar to the elite of the 19$^{\text{th}}$ century this group is overrepresented at the top but not underrepresented at the bottom. Interestingly, the size of the coefficients of the elite surnames that appear only in the pre-industrial period suggests that this group has been slowly converging to the mean. These findings corroborate that within the elite there are two different patterns of persistence. The first one corresponds to the elite surnames from the pre-industrial groups that were unable to enrol in the modern elite groups and shows that this group has slowly moved towards the mean status. The other pattern is the one of the long-surviving elites' surnames and the modern elite groups which indicates that these two groups have been more successful to maintain their social status and accessing high-quality education.

Table 5.6: The long-lasting elite

	Los Andes	International schools	High Quality Schools	HQPrivateSchool	Low Quality Schools	LQPublicSchool
Indigenous	-0.258^{***}	-0.124^{**}	-0.275^{***}	-0.222^{***}	-0.207^{**}	-0.131^{*}
	(0.056)	(0.040)	(0.054)	(0.044)	(0.075)	(0.064)
Afro-Colombians	-0.259^{***}	-0.132^{**}	-0.333^{***}	-0.271^{***}	0.578^{***}	0.238^{**}
	(0.066)	(0.048)	(0.050)	(0.047)	(0.147)	(0.089)
Multiperiod elite	1.208^{***}	1.053^{***}	0.811^{***}	0.998^{***}	-0.082	-0.010
	(0.247)	(0.206)	(0.141)	(0.189)	(0.090)	(0.074)
Only pre-industrial elite	0.142^{*}	0.162^{**}	0.211^{***}	0.206^{***}	0.030	0.100
	(0.069)	(0.049)	(0.057)	(0.050)	(0.080)	(0.067)
Only modern elite	1.634^{*}	1.375^{**}	0.846^{***}	1.139^{***}	-0.090	0.026
	(0.646)	(0.524)	(0.221)	(0.313)	(0.104)	(0.105)
Rare No Groups	-0.124^{*}	-0.015	-0.086	-0.058	-0.085	-0.009
	(0.054)	(0.035)	(0.047)	(0.039)	(0.068)	(0.054)
Constant	0.081	-0.016	0.050	0.022	0.072	-0.001
	(0.054)	(0.034)	(0.045)	(0.038)	(0.068)	(0.053)
Observations	9,922	9,922	9,922	9,922	9,922	9,922
Adjusted R^2	0.038	0.023	0.020	0.024	0.004	0.001

Note: Persistent elite is defined if a surname belongs to at least one pre-modern and one modern group. The equations are estimated using OLS.The 20 most commons surnames is the latent group of comparison.The RR_j is normalised. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

The institutional design of the Colonial system created segregated access to education, and as a consequence, elite–exclusive access. The Spanish Crown defined rules of access to higher education that forced the candidates to demonstrate their connection with family traditions linked to the traditional Spanish nobility. Towards the end of the colonial period, there was an elite of families of several generations born in the vice-royalty. These *Creole* elites were, in large part, protagonists of the process of independence and the creation of the Republic, at the beginning of the 19th century. Despite this political transformation, which sought to break with colonial forms, the rules of access to education did not transform substantially. Although the requirements for the purity of blood and nobility were relaxed, we show that *de jure* segregation in access to higher education continues.

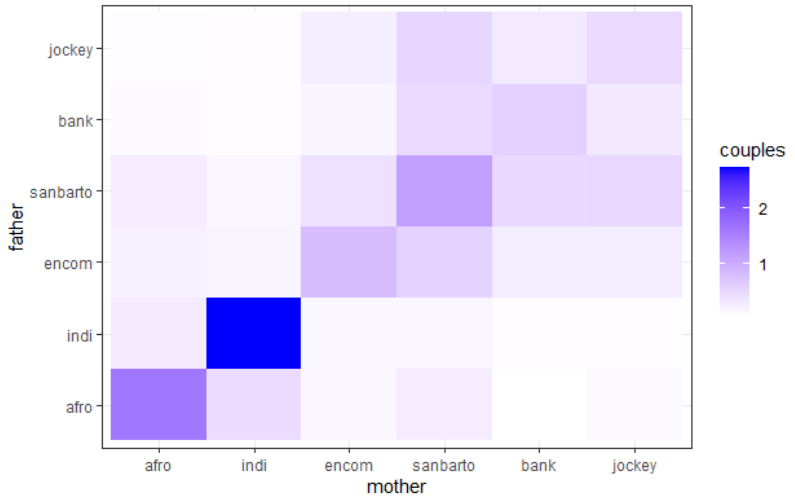
In conclusion, we find significant and robust differences in the relative representation between historical groups in access to education. These differences confirm the persistence of social segregation between ethnic and elite groups. Furthermore, a suggestive pattern of heterogeneity of these results among the elite groups emerges. First, the existence of convergence to the mean status of the pre-colonial elites. Second, the elites related to modern productive activities or the emerging bourgeoisie of the end of the 19^{th} century and the long-lasting elite present strong persistence of their privileged high status. Third, historically excluded social groups (African enslaved people and Indigenous) still show a lower social status in all the educational categories we use to measure it.

5.5.2 Assortative mating

Now, we turn our attention to the marriage patterns by looking at the proportion of paternal-maternal rare surnames pairs belonging to each historical group. Fig. 5.4 shows the results looking at the SIMAT sample. The diagonal of the matrix shows the proportion of intra-group marriage while the cells outside the matrix show the

inter-group matches. As expected, intra-group marriages are more prevalent and this is especially noticeable in the case of the ethnic groups that rarely match with other groups. In particular, ethnic groups seldom match elite groups of the late 19th century. On the contrary, elite groups from any period appear to match frequently with other elite groups. The results show that in the general population, there is high assortative mating in ethnic groups, little matching between ethnic and low homogamy across elite groups.

Figure 5.4: Who is marrying who: Population (SIMAT)



Notes: The results come from individuals that had two rare surnames identified in our historical groups. The total number of pairs found in each combination is normalised using the number of surnames in each list. Source: Authors' calculations with SIMAT.

To observe if homogamy reflects similar patterns of segregation as the ones shown in Fig. 5.3, we focus on three different social conditions: low-quality public schooling, high-quality private schooling and Los Andes University. When we compare the two most extreme social conditions, namely low-quality public and high-quality private schools, the results in Fig. 5.5a and Fig. 5.5b show that the marriage market is very different at these two points of the contemporary social ladder. The first thing worth noticing is that these results are in line with what we observed in Fig. 5.3, related to the relative presence of different historical groups. Afro-Colombians and

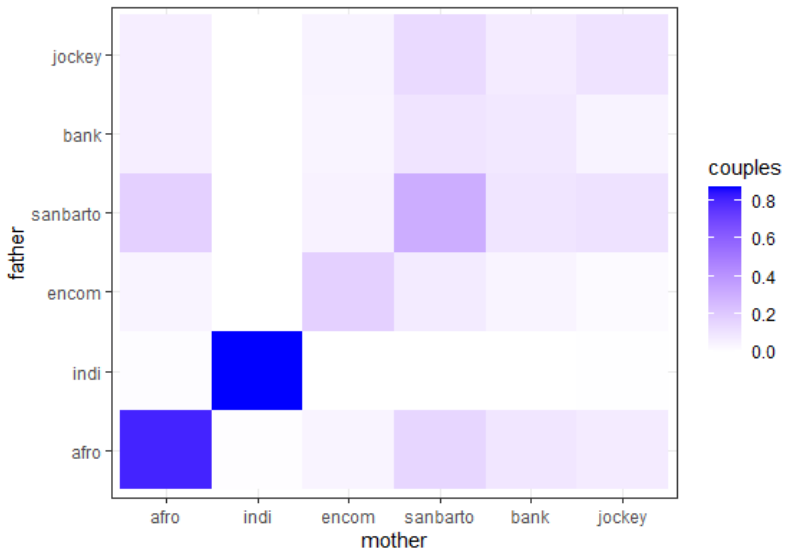
Indigenous appear more in low-quality public institutions, while they seldom appear in high-quality private institutions (the opposite is true for the elite groups).

In low-quality public schools, Afro-Colombians and elite surnames appear frequently matched to surnames from a distinct historical group, although in most cases it is more common to see a match between paternal and maternal surnames from the same group (as shown in the diagonal of the figure). On the contrary, Indigenous surnames show the highest incidence of homogamy as they only appear matched to other Indigenous surnames. The results are somehow similar to the marriage market in the general population (see Fig. 5.4), but there is more matching between Afro-Colombians and elite groups in low-quality public schools. Overall, we can conclude that low-quality public schools exhibit some degree of inter-group marriage between elite and Afro-Colombians, but none with Indigenous.

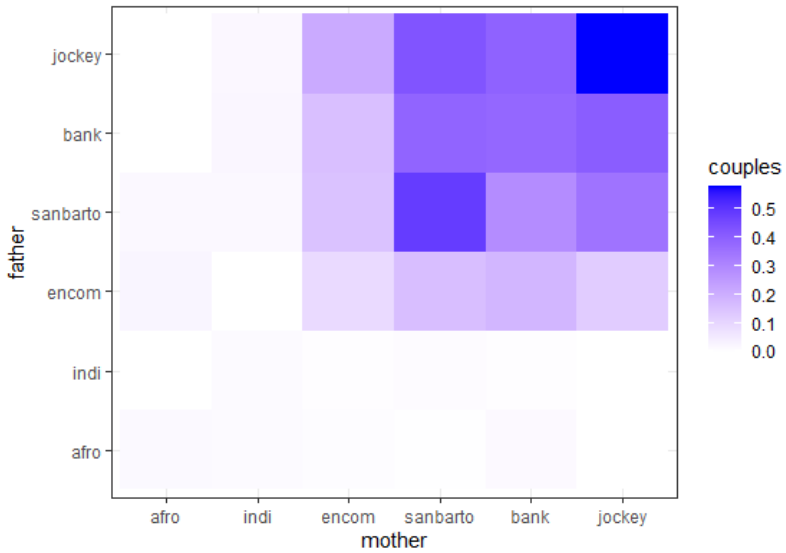
The marriage patterns clearly change when we look at the results in high-quality private schools. Maternal surnames from any elite group are rarely linked to Afrocolombian or Indigenous paternal surnames, and this is mainly because these groups are virtually absent in high-quality private schools. We observe a high frequency of marriage within elite groups which implies that the elite marriage market is very dynamic between elite groups. These figures reveal a striking difference between these two educational conditions. In low-quality public schools, there is more association across Afro-Colombians and elite groups, while in high-quality private schools there is clear segregation.

Figure 5.5: Who is marrying who: Public and private schools

(a) Low quality public schools



(b) High quality private schools

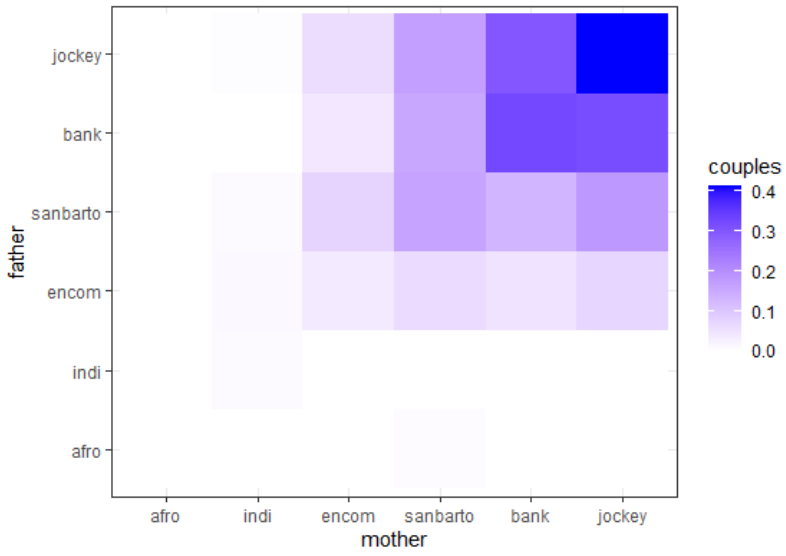


Notes: The results come from individuals that had two rare surnames identified in our historical groups. The total number of pairs found in each combination is normalised using the number of surnames in each list. Fig. 5.5a shows the marriage market in low quality public schools while Fig. 5.5b represents the marriage market in high quality private schools. Source: Authors' calculations with Saber 11.

When we repeat the exercise using graduates of Los Andes the results from Fig. 5.6 confirm that the same patterns are found in elite high schools and in elite higher education. Afro-Colombians and Indigenous surnames are practically absent

in Universidad de Los Andes, while maternal surnames from the elite are mostly found together with paternal surnames from the elite groups of the late 19th century. The results confirm that the segregation in the educational system is also reflected in segregated marriage markets.

Figure 5.6: Who is marrying who: High quality higher education



Notes: The results come from individuals that had two rare surnames identified in our historical groups. The total number of pairs found in each combination is normalised using the number of surnames in each list. Source: Authors' calculations with Los Andes graduates.

The differences are even more striking when we compare the total number of couples in the SIMAT and in each educational category. There are more couples formed by only Indigenous or only Afro-Colombians in SIMAT or in low-quality schools than in high-quality education (where they are virtually absent), while the number of elite couples is smaller in low-quality schools compared with SIMAT or high-quality schooling. The number of inter-group couples indicates that Indigenous rarely match with any other group. Out of 988,000 couples identified in SIMAT, there are around 160 couples formed by an Indigenous surname and an elite surname, or an AfroColombian surname and an elite surname while there are no couples formed by Afro-Colombians and Jockey surnames in high-quality private schools and in Los

Andes.

The previous results show that there is a marked trend toward homogamy among Indigenous and Afro-Colombians. Also, the different elite groups appear highly matched to members of the Jockey Club. This is particularly clear when we observe the patterns in high-quality private schools and universities. By contrast, and perhaps the most interesting result is the way in which most intergroup marriages appear only for children in low-quality public schools.

Table 5.7 reports the results of the Separable Extreme Value (SEV) index comparing different pairings of groups. Given that assortativeness is a local property the top part of the table reports the index between historical groups that are less likely to appear together, that is elite and ethnic. The second part of the table reports matching between our historical groups and the 20 most common surnames.

It is worth keeping in mind that when the SEV index is equal (or close to zero), there is random matching between the groups. Similarly, when the index is greater than zero, there is positive assortative matching or homogamy, that is that paternal surnames are only associated with maternal surnames of their own historical group. Thus, the SEV index confirms that in all social categories the level of assortativeness is always positive and there is little matching across elite groups and Indigenous and Afro-Colombians. Additionally, there is no matching between pre-industrial elite and modern elite groups with afro and Indigenous in Los Andes. The levels of the SEV prove that there is more mixing between Afrocolombian surnames and elite surnames in low-quality public schools than in high-quality public schools. The results corroborate that there is high homogamy in Indigenous groups across the different social categories as this group appears rarely matched to the other historical groups or common surnames, as the high SEV index indicates.

When we observe the matching of our social groups with common surnames

the SEV index reports that both pre-industrial and modern elite groups have an index close to zero, suggesting randomness in the matching between elite surnames and common across the different educational categories. This means that it is likely that elite surnames match with common surnames, in any of the educational conditions, based on the proportion of people holding an elite or a common surname. On the contrary, the Afro-Colombians and the Indigenous appear always with high levels of assortativeness even when compared with the common surnames and the SEV index is always higher than 2. Overall, we can conclude that there is virtually no mixing between elite and ethnic groups in Colombia in the educational system, not even in low-quality public schools.

Table 5.7: Separable Extreme Value index

	SIMAT	**Low quality public schools**	**High quality private schools**	**Los Andes**
pre-modern elite, Indigenous	3.95	8.05	3.64	-
pre-modern elite, afro	3.05	1.98	3.28	-
modern elite, Indigenous	4.75	-	5.02	-
modern elite, afro	4.17	3.23	-	-
Indigenous, afro	3.25	9.08	-	-
pre-modern elite, common	0.27	0.55	0.23	0.44
modern elite, common	0.52	0.75	0.58	1.04
afro, common	3.75	2.63	3.52	-
Indigenous, common	3.99	7.38	2.28	3.53

Note: The table shows the results from Eq. (5.3). When matching is random the index will be equal to zero and if there is positive assortative matching then the index will be greater than zero.

These results present novel empirical evidence of the presence of assortative matching in the contemporary marriage market. In particular, these findings suggest that the high levels of homogamy reinforce the persistent segregation of the educational system in contemporary Colombia, especially for ethnic surnames.

5.5.3 Caveats

Although the results consistently showed that segregation in the educational system today has its roots in the deep past, there are several concerns to consider. The main concern is that our sample of rare surnames is defined using a contemporary source of surnames. If certain historical groups were more likely to become extinct, our estimations would be biased given that we are selecting a sample of families that behave differently from the population. The last column of Table 5.8 shows the percentage of extinguished surnames, or those that appear in the historical record but not in the paternal surname in SIMAT. This caveat seems to be especially important for the encomenderos and the Afro–Colombians.

Table 5.8: Linking historical groups with contemporary surnames

Historical Group	Number of surnames	Num. surnames in SIMAT	Percentage of non matched
Indigenous	131	113	11%
Afro–Colombian	271	65	66%
Encomenderos	363	55	46%
slaves owners	1,607	578	33%
Bank-shareholders	241	66	12%
Jockey Club	309	97	12%
San Bartolome	224	224	34%
U Rosario	711	238	24%

Notes: The first column shows the number of surnames in each list of historical groups. The second column is the number of surnames we define as rare using the number of holders in SIMAT. The last column shows the percentage of extinguished surnames, or those that appear in the historical record but not in SIMAT.

Following genealogical trees from different sources, we have seen that in the case of the encomenderos the transmission of surnames between parents and children has not always been through the paternal and maternal first surnames, as is the case today. Up until the end of the 18^{th} century, surnames could vary among members of the same family, weakening the link between the list of encomenderos and contemporary status measures. Additionally, compound surnames have been transformed and simplified over time. An example of this is the surnames Arias de Monroy or Enciso y Cardenas. The addition of the particle "y" used to be customary

to avoid confusion between given names and surnames.[23] Therefore, it is likely that surnames before the late 18th century are less informative and that the links between generations are weaker.

In the case of the Afro–Colombians, the sources that we use to compile this list do not provide detailed information about the number of holders of the surnames at the time of compilation. Furthermore, part of the surnames come from registers of slave importation, or African enslaved people that entered the port of Cartagena, which means that it is possible that some surnames from the list did not stay in Colombia.[24]

A similar concern is that by looking only at contemporary surnames, we select surnames that were common in the past and only recently became rare. To minimise this concern, we consider several definitions of rarity and report the results in the Appendix. Although we do not have reasons to believe that the distribution of surnames has changed differently across our historical groups, in the future we will do a cohort-based analysis, looking at the distribution of surnames by cohorts using the 2018 census data. Additionally, we should highlight that the common surnames that we use as our comparison group correspond to surnames that have been traditionally common even in Spain. The 20 surnames that belong to our commons group end in *-ez* which meant *"son of"*, similar to the ending *-son* in English surnames.

Another potential concern regarding the interpretation of the slaves owners' regression to the mean is that this could be explained by the presence of surnames that also appear in Afro-Colombians. To alleviate this concern, we calculate the

[23]In Spanish, it is common to find surnames, like Ramon, that are also first names. In this case, surnames are separated by the particle as in Rodrigo Ramon y Casales. This rule was also adopted if both surnames have the particle "de", to avoid confusion and to clearly differentiate both surnames, as in Francisco de Goya y Luciente.

[24]We also checked with Census data and found 152 surnames in the 2018 census. Checking the origin of some of the surnames that do not match, we found that several of these surnames appear in other Latin American countries but not in Colombia.

average relative representation of all surnames from the slaves owners list excluding those that appear with at least one incidence in the 2010 and 2013 waves of the longitudinal survey, ELCA.[25] The results in Fig. 5.7 show very similar patterns that the original results with all the slaves owner's surnames. Overall this group appears overrepresented in both low and high-quality education, indicating that the regression to the mean of this group is not driven by the presence of Afro-Colombian surnames.

Figure 5.7: Relative Representation in access to education - slaves owners

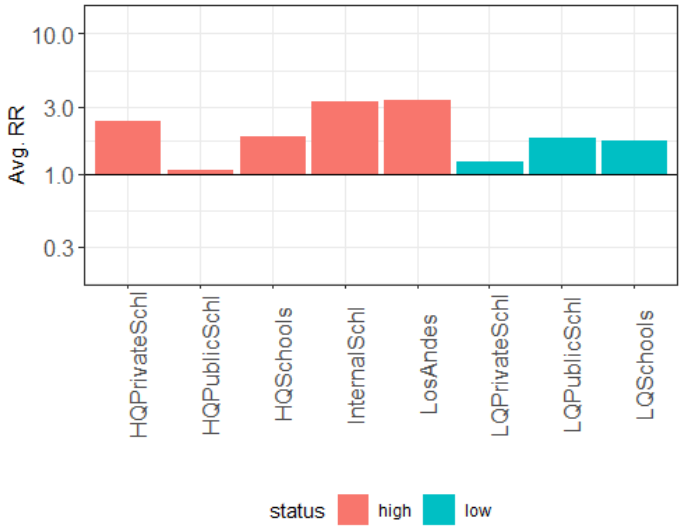


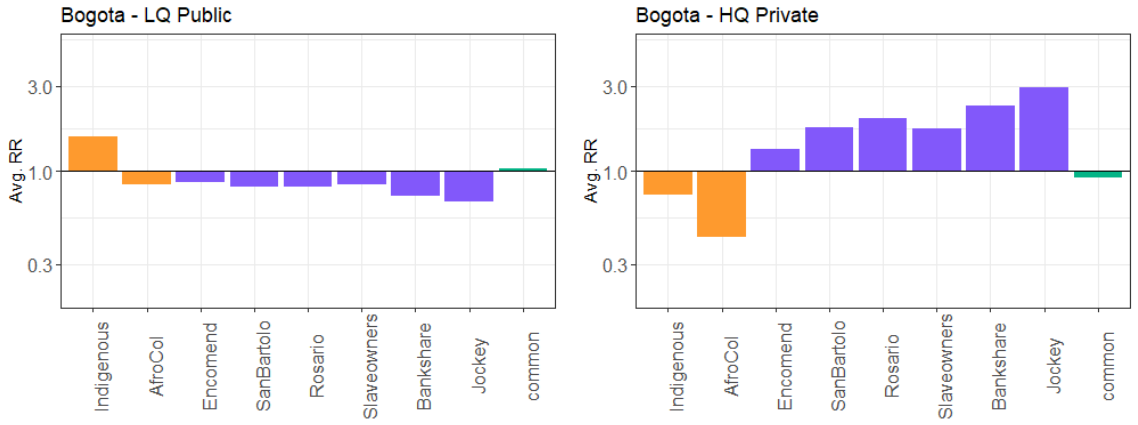
Fig. 5.2 presents the average relative representation of the 20 most common surnames in Colombia.The horizontal black line is set at $RR_j = 1$, corresponding to the neutral level of relative representation.

Although the findings show that historically marginalised groups are excluded from high quality education, it is possible that Indigenous and Afro-Colombian surnames come disproportionately from unusual and/or isolated populations and therefore we are observing two groups that are different from the rest. To address this, in Fig. 5.8a and Fig. 5.8b we show the distribution of the relative representation of each historical group in low and high quality schools in Bogota. Looking at Bogota the results show that even in Bogota, Indigenous and Afro-Colombians are excluded from High-quality private schools and appear slightly overrepresented in Low-quality

[25]This survey follows approximately 10,000 Colombian households in urban and rural areas of Colombia every three years and reports among other characteristics, ethnicity.

public schools. These results also confirm that even in the capital city, where the supply of education both private and public, is high, ruling out that the results are only driven by the lack of supply of education.

Figure 5.8: Relative Representation in schools in Bogota



(a) Low quality public schools (b) High quality private schools

Notes: The figure shows the distribution of the relative representation by historical group in the access to schools in Bogota. Source: Authors' calculations with Saber 11 and SIMAT.

Finally, we have previously discussed that there is a potential measurement error in the status of certain surnames as we are assuming that all groups are homogeneous in status. This seems to be the case in particular with encomenderos and slaves owners. However, as shown in Table 5.9, the number of elite families from the encomenderos and slaves owners that are found in both the pre-industrial and the industrial elites is lower in comparison to the number of surnames of the colonial schools. This suggests that both encomenderos and an important part of the slaves owners families were less likely to belong to the modern elite, which supports the hypothesis that pre-industrial economic activities based on slave labour were detrimental to maintaining high status in education.

Table 5.9: Number of families by period

	Only pre-industrial elite	Only modern elite	Multi-period elite
Encomendero	58	-	8
slaves owners	585	-	87
Urosario	206	-	68
SanBartolome	181	-	49
Bankshareholders	-	17	62
JockeyClub	-	29	88

5.6 Conclusions

This chapter brings novel evidence that shows that social segregation is present in the contemporary educational system in Colombia. We confirm that this segregation reproduces patterns of social exclusion that go as far as the Colonial period, for example, the Spanish Crown defined rules of access to higher education that forced the candidates to demonstrate their connection with family traditions linked to the Spanish nobility. Although the requirements for the purity of blood and nobility were relaxed after the Independence, in the early-19$^{\text{th}}$ century, we show that *de jure* segregation in access to higher education continues.

In this chapter, we assume that rare surnames can be treated as belonging to the same extended family, which allows us to establish links between multiple–generations. Using historical primary sources and ethnographic studies, we identified surnames belonging to elites and ethnic groups throughout history. We find very sticky floors for the ethnic groups, while the elite of the past has been slowly moving down the social ladder. In particular, the persistent under-representation of historically excluded social groups (Afro-Colombians and Indigenous) in high–quality education is noticeable. Additionally, we show that there is high assortative matching in the marriage market, particularly in the case of ethnic groups, and this reflects social segregation at the school and university levels. We conclude that there is no

matching in the marriage market between elite and ethnic groups in Colombia.

We find three different mobility patterns in the history of Colombia. First, pre-industrial elite surnames have slowly moved towards the mean status, as these surnames do not appear over or underrepresented in almost any educational category. Second, elite groups from the late 19^{th} century and surnames that appear in both periods show considerable persistence in their social status, appearing at the top of the social ladder in several markers of status. The elite could move down the social ladder, but this is a very slow process. Third, ethnic groups have the highest persistence of social status among the groups we studied. While we observe some reinvention of the elite, the bottom part of the social status has continued trapped with very scarce possibilities to move upward.

These results present novel empirical evidence of the presence of assortative matching in the contemporary marriage market. In particular, our findings suggest that the high levels of homogamy reinforce the persistent segregation of the educational system in contemporary Colombia, especially for ethnic surnames.

While previous literature has shown that extractive institutions such as the *encomienda* and slavery had different long-term effects at the municipal level, we find that the effects were uneven for different social groups. Although the *encomienda* had long-term positive effects on development, it negatively affected the Indigenous population. On the contrary, the social status of the encomenderos has converged to the mean, especially for those families that were unsuccessful in enrolling in new modern activities. Similarly, the literature has shown that slavery had detrimental effects on several socioeconomic outcomes at the municipal level. Our results uncover that these effects are especially persistent for Afro-Colombians, while the social status of slaves owners has regressed to the mean, as in the case of the encomenderos. These results show that the Indigenous and Afro-Colombians, which were at the bottom of

the social ladder during the colonial period, appear at the bottom of the contemporary educational system. The elite status of the encomenderos and slaves owners, on the contrary, has not persisted over time. However, the social status of elite groups of the late 19^{th} century and the status of elite families that persisted from the colonial period to the modern one strongly predicts the contemporary social status today.

Furthermore, our results can also shed light on the effects of these institutions on inequality. The literature suggests that the *encomienda* has little effect on contemporary inequality, while the presence of slavery is strongly associated with land inequality today. It is possible that as the Indigenous were displaced from their land with the collapse of the encomienda, the negative effects of the *encomienda* on these groups are not captured at the municipal level. On the other hand, after the prohibition of slavery, the Afro-Colombians stayed in the same zones, reinforcing the negative effects of slavery.

5.A Appendices

5.A.1 Description of contemporary educational data

To define contemporary access to education we construct several educational categories (e.g. high and low-quality education, high prestige high schools and higher education institutions). We provide a short description of each dataset and the variables we use to define the social conditions under consideration in our analysis of the persistence of social status.

The first dataset is an administrative register of the SABER 11. It contains information at the individual level compiled from the registration forms filled out by students who finish high school and take a standardised test, known as SABER 11, similar to the SAT in the USA or the A-levels in Europe. This test is a control of

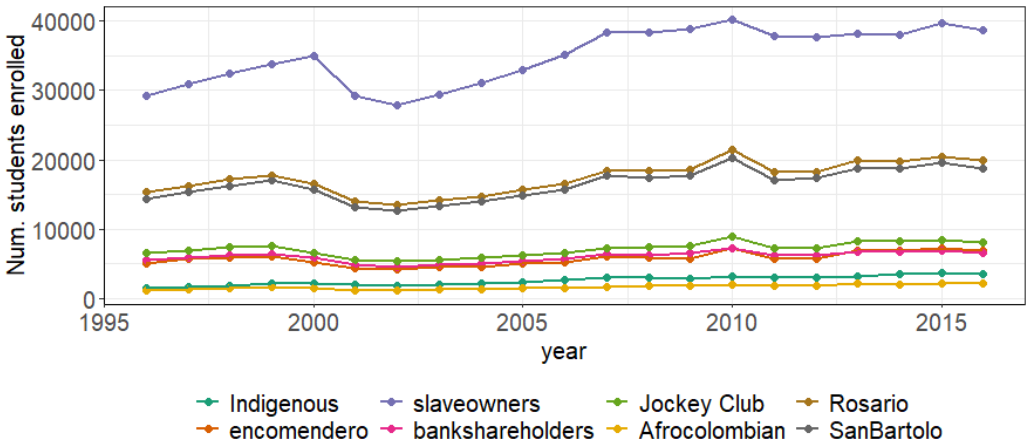
the quality of the educational attainments of students and serves also as a college admission test. It is administered and applied by the ICFES (Instituto Colombiano para el fomento de la educación superior). We use information from these records for the years 1996 to 2016 on an annual basis. The database contains a total of 12,456,376 individuals. Fig. 5.9 shows the number of students enrolled in the last year of secondary school by the historical group from 1996 to 2016. There are no considerable differences in the trends of the historical groups, but the differences in levels are significant. The ratio between the number of students in the third grade of primary school and the number of students enrolled on the last year of secondary schooling in 2016 is around 35%, as shown in Table 5.10.

Table 5.10: Completion rates by historical groups.

Historical group	Num students enrolled SIMAT, 2016	Num students enrolled SABER 11, 2016	Saber11/SIMAT (%)
Indigenous	10,685	3,577	33.48
Afro–Colombian	6,816	2,097	30.77
Encomenderos	17,487	6,943	39.70
slaves owners	108,386	38,606	35.62
San Bartolome	50,471	18,700	37.05
Rosario	53,641	19,877	37.06
Bankshareholders	16,711	6,618	39.60
Jockey Club	21,426	8,086	37.74
Common surnames	395,654	134,203	33.92

The second column shows the number of students enrolled in third grade of primary schooling in 2016 by historical groups. The third column is the number of the number of students enrolled in the last year of high school in 2016 by historical groups. The last column is the proportion between column 2 and 3 as a way to calculating the share of students in high school in comparison to primary schooling.

Figure 5.9: Students enrolled in high school, 1996-2016



The figure shows the number of students enrolled in the last year of high school by historical group using rare surnames, from 1996 to 2006. Source: SABER 11.

We use the school-level scores in the standardised tests to define High or Low-quality schools. We label these social conditions as High or Low-Quality Schools in general (HQRanking, LQRanking) or High-Quality Private or Public Schools (HQPrivateSchool, HQPublicSchool).[26] The distinction between private and public is crucial in a country like Colombia where private schooling covers a significant proportion of the schooling supply.

Besides the high-school institutions' quality define above, we also use the dataset SABER 11 to identify a marker of an elite education: the UNCOLI schools, for Union de Colegios Internacionales (International High–Schools Union). This union groups the most prestigious, selective and expensive schools in the country, located in Bogota.

Finally, Los Andes is a private institution founded in Bogotá in 1948 and was established to be independent of any political party or religious institution. Since then it has been considered a particularly elite institution. Since 1950, less than 0.03% of people between 20 and 30 years have graduated from this university. Of the 363,039 students who registered in a university in 2016, 2,940 enrolled in Los Andes, that is, only 0.8%. Nevertheless, the last government was composed of 33% of graduates from Los Andes. Graduates from this university are the most represented in high-level policy-making. Also, alumni of Uniandes (as known in Colombia) are over-represented among the private managers of the biggest companies in the country.

[26]It is important to note that a significant part of the population does not belong to any of those extreme social conditions. We choose to create such an exclusive set of social conditions because we are aiming at understanding the persistence and the mobility of elites and ethnic groups along the social ladder.

5.A.2 Description of relative representation by educational outcomes

Table 5.11: Summary statistics of the relative representation by historical groups

Educational category	Min.	1stQu.	Median	Mean	3rdQu.	Max.
Uniandes	0	0	0	1.12	0.65	121.09
Uncoli - International schools	0	0	0	1.16	0	211.28
HQ Universities	0	0.18	0.61	0.83	1.14	29.57
HQ Ranking	0	0	0.36	0.99	1.16	46.46
HQ Public school	0	0.09	0.52	0.91	1.10	40.47
HQ Private school	0	0	0	1.03	0.95	64.06
LQ Ranking	0	0	0.15	1.21	1.37	164.61
LQ Public school	0	0	0	1.23	1.19	191.20
LQ Private school	0	0	0	0.86	0.97	54.54

Source: Authors' calculations with Los Andes graduates, SIMAT, and SABER 11.

5.A.3 Robustness: Different thresholds of rarity

To address potential selection in our rare surnames we consider other definitions of rarity. First, we include all the surnames. The results show that the general findings hold, but, as expected, the significance and explanatory power of the regressions fall as the predictive power of frequent surnames is lower than the predictive power of rare surnames. Then, we impose a more restrictive definition of rarity (1 % of the most common surname). The results reveal that the patterns remain but the significance of every coefficient improves as well as the overall fit of the regression.

Table 5.12: All surnames. Differences in educational social categories - OLS z-scores

	High Quality Schools	Low Quality Schools	LQPrivateSchool	LQPublicSchool	HQPrivateSchool	HQPublicSchool	Los Andes	International schools
Ethnic groups								
Indigenous	−0.618***	−0.142	−0.232**	−0.121	−0.578***	0.057	−0.632***	−0.511***
	(0.106)	(0.106)	(0.106)	(0.106)	(0.106)	(0.106)	(0.105)	(0.106)
Afro-Colombians	−0.584***	0.352***	0.102	0.095	−0.519***	−0.219*	−0.558***	−0.452***
	(0.112)	(0.113)	(0.113)	(0.113)	(0.112)	(0.113)	(0.112)	(0.112)
Pre-industrial elites								
Encomendero	−0.301***	−0.046	−0.039	−0.027	−0.344***	−0.026	−0.383***	−0.297***
	(0.080)	(0.081)	(0.081)	(0.081)	(0.080)	(0.080)	(0.080)	(0.080)
slaves owners	−0.258***	0.075	0.036	0.078	−0.257***	−0.050	−0.390***	−0.332***
	(0.055)	(0.055)	(0.055)	(0.056)	(0.055)	(0.055)	(0.055)	(0.055)
URosario	−0.029	−0.017	0.002	−0.017	−0.115**	0.052	0.008	−0.099*
	(0.056)	(0.056)	(0.056)	(0.056)	(0.056)	(0.056)	(0.056)	(0.056)
SanBartolome	−0.212***	−0.127**	−0.038	−0.116**	−0.257***	−0.0005	−0.401***	−0.259***
	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)
modern elites (late 19th century)								
Bankshareholders	0.121	−0.081	0.068	−0.063	0.234***	0.029	0.097	0.004
	(0.082)	(0.082)	(0.082)	(0.082)	(0.081)	(0.082)	(0.081)	(0.082)
JockeyClub	0.503***	−0.053	−0.103	−0.057	0.660***	0.041	0.831***	0.780***
	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)
Rare No Groups	−0.492***	−0.107*	−0.157***	−0.087	−0.449***	−0.256***	−0.538***	−0.424***
	(0.054)	(0.055)	(0.055)	(0.055)	(0.054)	(0.055)	(0.054)	(0.054)
Constant	0.473***	0.099*	0.145***	0.081	0.433***	0.238***	0.521***	0.412***
	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)
Observations	23,144	23,144	23,144	23,144	23,144	23,144	23,144	23,144
Adjusted R^2	0.010	0.002	0.002	0.001	0.010	0.004	0.014	0.009

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: The equations are estimated using OLS.
The 20 most commons surnames is the latent group of comparison.
The RR_j is normalized .

Table 5.13: Rare surnames : 1% of Rodriguez. Differences in educational social categories - OLS z-scores

	High Quality Schools	Low Quality Schools	LQPrivateSchool	LQPublicSchool	HQPrivateSchool	HQPublicSchool	Los Andes	International schools
Ethnic groups								
Indigenous	−0.456***	−0.197	−0.293**	−0.178	−0.384***	0.118	−0.484***	−0.325***
	(0.122)	(0.123)	(0.123)	(0.123)	(0.121)	(0.123)	(0.120)	(0.121)
Afro-Colombians	−0.496***	0.565***	0.207	0.173	−0.417***	−0.146	−0.468***	−0.317**
	(0.137)	(0.139)	(0.139)	(0.139)	(0.137)	(0.139)	(0.135)	(0.137)
Pre-industrial elites								
Encomendero	−0.122	0.102	0.116	0.143	−0.223	0.033	−0.313**	−0.224
	(0.161)	(0.162)	(0.163)	(0.163)	(0.160)	(0.163)	(0.159)	(0.160)
slaves owners	0.039	0.114	0.014	0.125	0.095	0.001	−0.045	−0.014
	(0.078)	(0.078)	(0.078)	(0.078)	(0.077)	(0.078)	(0.076)	(0.077)
URosario	0.208***	0.024	0.078	0.017	0.218***	0.078	0.325***	0.263***
	(0.080)	(0.081)	(0.081)	(0.081)	(0.080)	(0.081)	(0.079)	(0.080)
SanBartolome	−0.010	−0.205**	−0.086	−0.200**	−0.073	0.062	−0.293***	−0.202**
	(0.088)	(0.089)	(0.089)	(0.089)	(0.088)	(0.089)	(0.087)	(0.088)
modern elites (late 19th century)								
Bankshareholders	0.381***	−0.061	0.094	−0.046	0.557***	0.020	0.117	−0.101
	(0.139)	(0.140)	(0.140)	(0.140)	(0.138)	(0.140)	(0.137)	(0.138)
JockeyClub	0.663***	−0.081	−0.111	−0.066	0.873***	−0.078	1.679***	1.536***
	(0.113)	(0.114)	(0.114)	(0.114)	(0.113)	(0.114)	(0.112)	(0.113)
No Group	−0.276***	−0.082	−0.158**	−0.062	−0.226***	−0.111	−0.353***	−0.219***
	(0.078)	(0.078)	(0.079)	(0.079)	(0.078)	(0.079)	(0.077)	(0.077)
Constant	0.240***	0.069	0.142*	0.053	0.189**	0.096	0.309***	0.186**
	(0.077)	(0.078)	(0.078)	(0.078)	(0.077)	(0.078)	(0.076)	(0.077)
Observations	9,614	9,614	9,614	9,614	9,614	9,614	9,614	9,614
Adjusted R^2	0.023	0.004	0.003	0.002	0.029	0.001	0.053	0.035

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: The equations are estimated using OLS.
The 20 most commons surnames is the latent group of comparison.
The RR_j is normalized .

6

Social mobility: the case of Los Andes University, 1949-2018.

6.1 Introduction

Social mobility, defined as the easy upward and downward movement of generations along social status, has been widely studied in the literature, often using surveys available from the 1960s onward.[1] However, historical social mobility has been more difficult to study as measuring this process requires significant amounts of longitudinal data.[2] With the digitisation of full count historical censuses and the methods of automated record linkage, the study of social mobility has gone as far back as the last two centuries. The long-run mobility estimates for the United States (Olivetti & Paserman, 2015; Feigenbaum, 2018; Ward, 2021), England (Long & Ferrie, 2018; Pérez, 2019), Sweden (Berger, Engzell, Eriksson & Molinder, 2021) or Argentina (Pérez, 2017) go as back as the mid-19$^{\text{th}}$ century.

Nonetheless, there is no consensus regarding over how many generations status persists or even more crucially how to measure accurately social status (Clark, 2007; Vosters, 2018; Solon, 2018). Social status has been measured using literacy (Beltran Tapia & de Miguel Salanova, 2021), occupation (Berger et al., 2021; Long & Ferrie, 2018; Pérez, 2019) or earnings (Feigenbaum, 2018). The ideal measure should show how social status transmits across generations, but education, occupation, or permanent income are not always observed at the same time in the generations studied. Also, the use of single observations of social status results in measurement error and attenuation bias in the intergenerational coefficient (Ward, 2021; Santavirta & Stuhler, 2020; Clark & Cummins, 2013; Clark et al., 2015). Another important limitation when comparing different measures of social mobility is how it is estimated. Some have focused on estimating an intergenerational coefficient of status (Clark et

[1] There is a considerable amount of literature on social mobility both empirical and theoretical. For a complete literature review see Black and Devereux (2011) and Solon (2018).

[2] Recently, Guell et al. (2013) proposed a different method for measuring intergenerational mobility that requires a single large cross section reporting both surnames and a measure of social status: the "Informative Content of Surnames" (ICS). See Barone and Mocetti (2021) and Gabbuti (2022) for an application of the ICS in a historical setting.

al., 2015; Ward, 2021) while others have used measures such as the Altham statistic to measure occupational mobility (Long & Ferrie, 2013; Modalsli, 2015; Pérez, 2017; Cilliers & Fourie, 2018).

Facing similar limitations the research on social mobility in Colombia has focused on the second half of the 20th century using survey data with retrospective questions (Angulo et al., 2012; Ramirez-Zuluaga, 2016; Moyano & Galvis, 2014; Galvis & Meisel, 2014; Dahan & Gaviria, 1999). In general, the studies find that social mobility accelerated especially during the last three decades of the 20th century and slowed down and slightly decreased during the last ten years.[3] But despite the recent improvement in the different indicators of social mobility, Colombia is still considered one of the countries in the world with lower social mobility, where "moving from rags to riches in one generation is virtually impossible" (Behrman et al., 2001, p.11).[4] Additionally, Ramirez-Zuluaga (2016) estimates that the children of a low-income family would require more than 10 generations to reach the mean (see also OECD, 2018, p.27). Social mobility figures for the 19th or early 20th century are still missing from the literature and this chapter aims at providing a long-term perspective of social mobility building upon the findings of Chapter 5.

As social groups exhibit different patterns of status persistence, we can also expect different patterns of social mobility in the long run. I follow dynasties of rare surnames as proposed by Guell et al. (2007), Clark (2014), and others.[5] Assuming that rare surnames are part of the same extended family I can trace families from colonial times, and I can follow their social status over generations. Following Chapter 5, I use administrative data from the register of enrolment in primary education SIMAT

[3]One exception is the work of Woessmann, Ursprung and Schuetz (2005) that concludes that in Colombia, family characteristics have a relatively minor effect on the educational outcomes of the student.

[4]See also Narayan et al., 2018 and OECD, 2018.

[5]See Clark et al. (2015) for an overview of the methodology and Santavirta and Stuhler (2020) for a literature review on the methods.

(for the name in Spanish: "Sistema de Información de Matrícula") to have the most complete sample of Colombian surnames. Also, I use the data described in Section 5.3 to define social status attributes for different groups, at different points in history.

The closest example of an application of this method to a society similar to Colombia is Diaz Vidal (2014). The author calculates the inter-generational correlation of occupational income for Chile. He studies two contemporary (late-20th century) generations in Chile using 14 historical groups of surnames and focuses on a comparison between Chilean Mapuches and elite groups (landowners of the 19th century and European migrants). Similar to Diaz Vidal (2014), my study covers two historically marginalised groups (Afro-descendants and pre-Hispanic Indigenous groups) and a wider set of elite groups. For the latter, I consider elite groups that go as far back as the colonial period allowing for a better understanding of the long-term process of social status in their institutions.

As I do not have a measure of income at the surname level across generations, I estimate the relative representation of surnames in the graduates of an elite university from 1949 to 2018. Los Andes University is a private university founded in Bogota in 1948 and it is regarded as one of the best, and most exclusive, universities in the country.

I divide the sample of graduates into two cohorts of 35 years: those graduating between 1949-1983 and those graduating between 1984 - 2018.[6] Comparing the relative representation in both cohorts, I provide a lower bound of the intergenerational correlation of status, at the surname and the group level. At the surname level, I estimate a coefficient of intergenerational mobility of 0.52, which is in line with other estimations and suggest slow regression to the mean. At the group level, my results

[6]I do not have data on the age of graduation of each individual but it is reasonable to assume that people graduate from university when they are around 25 to 35 years old. If this is the case, my estimations compare a generation of students born between 1915 to 1949 and a generation born between 1950-1993.

show that historical groups that represented the elite in the sixteenth, nineteenth and twentieth centuries appear overrepresented in both cohorts of graduates of this university. Historically marginalised groups appear always underrepresented in Los Andes.

The findings of this chapter confirm the low-mobility patterns found in other studies for Colombia. Families that were over-represented in Los Andes in the past, will be over-represented in the future. But the main contribution is to show that the slow regression to the mean is especially clear for the elite groups that appear both in pre-modern groups and in the late 19^{th} and beginning of the 20^{th} centuries. For these families, I find a coefficient of intergenerational mobility of 0.55. The under-represented groups have a small and non-significant correlation coefficient, explained mainly by their persistent exclusion during the whole period. Finally, although the elite groups of the Colonial period have a relative representation in Los Andes that is close to one, these groups have not completely regressed to the mean, indicating low mobility, even for an elite group of the 17^{th} century. For colonial groups, I find an intergenerational coefficient between 0.44 to 0.61. The results are consistent with the findings of Chapter 5 that reveal different patterns of persistence and social mobility by different historical groups. Similar to other research for Colombia, the findings of this Chapter show that social mobility is very slow in the upper and lower parts of the social status ladder (Angulo et al., 2012; Behrman et al., 2001; Ramirez-Zuluaga, 2016; Hertz et al., 2007).

This research shows that the application of this type of method can be especially useful to study countries like Colombia, where there is no consistent and long-term information to calculate social mobility. The use of surnames seems particularly relevant in this context where the work of professional and amateur genealogists, who have contributed to an impressive amount of data, remains unexploited

by quantitative analysis and economic historians in general (Clark et al., 2015; Mejia, 2012).

6.2 Data

The study of very long–term, multigenerational social mobility has gathered interest in different fields of social sciences but requires defining historically consistent metrics to track status over time. Therefore, the most important empirical contributions come from the examination of social mobility using huge databases containing administrative registers and longitudinal individual data (see for example Chetty et al., 2014, Abramitzky et al., 2020, Ward, 2021). However, the study of historical social mobility is hindered not only by the availability of long longitudinal series, linking the same family through time but also by the stability of the measurement of socio-economic performance that one wishes to measure.[7]

To measure social status in several generations I use the graduates from Los Andes University and observe how access to elite tertiary education has changed over time.

Los Andes University was founded on the 16^{th} of November 1948 by a group of intellectuals and entrepreneurs led by Mario Laserna Pinzon, with the main objective of creating an independent, non-religious non-politically affiliated university to train competitive professionals with the highest international standards. The foundation of the university also responded to the new liberal Governments of the 30s and 40s, which started the process of mass education in the country with the idea of modernising the country through education. The university opened its doors offering 7 degrees to 79 students and started with 16 professors, but the numbers grew steadily after that. The first graduates from the University were economists and soon after a

[7]For different measures of social mobility see Chetty et al. (2014) and Vosters (2018).

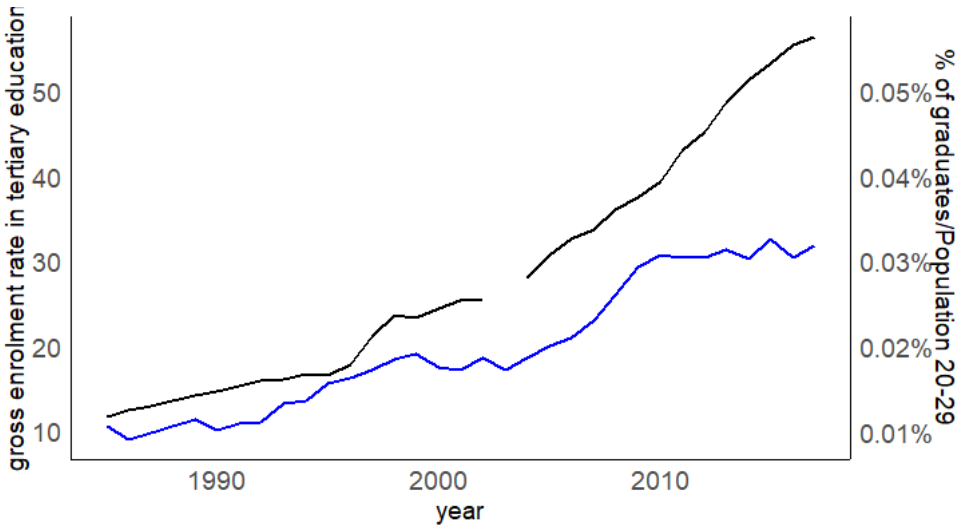
cohort of engineers and architects graduated in 1954.

From 1950 to 2016, Colombian governments were mostly composed of graduates from the Los Andes University. Likewise, the fiscal and monetary policy of the country was directed by graduates of this university: more than 60% of the directors of the Central Bank and the Department of National Planning had an undergraduate degree from Los Andes (Pearce & Montoya, 2022). Graduates from Los Andes also have striking advantages in the labour market. According to a recent report from the University, 95% of its graduates are employed and earn an average salary of 2,000 USD per month (Universidad de los Andes, 2022). In comparison, the national youth unemployment rate (people between 15 and 28 years old) is 18% and the average monthly salary is 300 USD.

The literature has pointed out that education has in itself the capacity to promote social mobility, and by expanding the provision of education, a democratising process of occupational opportunities can be generated (Breen & Karlson, 2014; Torche, 2014). However, as argued by Chetty et al. (2017) and Cardenas et al. (2021), segregation in access to high-quality education can limit social mobility.

This could be particularly important in a country like Colombia which witnessed an increase in tertiary education from 4% in 1970 to 60% by 2017, according to the World Bank. However, since the 1980s less than 0.03% of people between 20 and 29 years have graduated from this university (as shown in Fig. 6.1), and of the 363,039 students who registered in a university in 2016, 2,940 enrolled in Los Andes, which is the equivalent to only 0.8% of students registered.

Figure 6.1: Enrolment in tertiary education and graduates of Los Andes



Notes: The black line shows the gross enrolment rate in tertiary education from the World Bank. The blue line shows the Graduates of Los Andes as a proportion of the total population at ages 20-29, which correspond to the average ages of people graduating from the university. Although the proportion has increased in the last 30 years, the proportion is still less than 0.03% Source: The number of people age 20-29 comes from Department of National Statistics (DANE) and the number of graduates comes from Los Andes.

To track generations over time, I use administrative data from the register of enrolment in primary education SIMAT to have the most complete sample of Colombian surnames to define rare surnames that I can track back to the Colonial period. Following the previous chapter, my preferred definition of rare surnames is those that have less than 2% of the number of holders of the most common surname in the sample (Rodriguez) and those with at least 3 holders (to avoid possible spelling errors). I modify the definition of rarity using different thresholds but overall I do not find significant differences explained by the threshold. Then, I use the data described in Section 5.3 to assign to the rare surnames a historical social status. This dataset provides lists of surnames that belong to a specific social group that can be related to a specific period in the history of Colombia. In total, I analyse six elite groups and two ethnic groups: *encomenderos* (Spaniard colonial officers) of the late 16^{th} century and slaves owners in 1851, represent the colonial elite, students of the first schools founded in Bogota, founders of banks during the free-banking system in

the 1870s and the founders of the first social club in Colombia established in 1874 represent the beginning of "modern elite". It should be also noticed that Indigenous and Afro-Colombian surnames represent social groups that have been historically marginalised and as this research will confirm, have been historically absent from Los Andes University.

Table 6.1 summarises these groups and Section 5.3 presents a detailed description of each group in the historical context.

Table 6.1: Historical groups

Social condition	Historical group	Period	Characteristic
Ethnic groups	Indigenous	Forced work since 15^{th} century	racial/ethnic group
	Afrocolombian	Forced work since 16^{th} century	racial/ethnic group
Pre-industrial elite	Encomenderos	16^{th} – 17^{th} century	Colonial elite
	Slave owners	Received compensation from the state in 1851	Colonial elite
	Colegio San Bartolome	1605 – 1820	Educational elite
	Colegio del Rosario	1773 – 1842	Educational elite
Modern Elite	Bank founders	1870 – 1885	Business elite
	Jockey Club	1874 – 1902	Social elite

6.3 Historical groups in Los Andes University

The presence of each historical group in Los Andes from 1949 to 2018 is summarised in Table 6.2 and Fig. 6.2. Table 6.2 shows the total number of surnames compiled for each historical group, the number of rare surnames that matched SIMAT using the preferred definition of rarity, and the total number of graduates in Los Andes from 1949 to 2018. While for elite groups the number of graduates in Los Andes is more than 10 times the number of surnames found in SIMAT, for the Indigenous and Afro-Colombians, there are half of the graduates in Los Andes as surnames in SIMAT. This indicates how rare is to find an ethnic surname among the graduates of Los Andes.

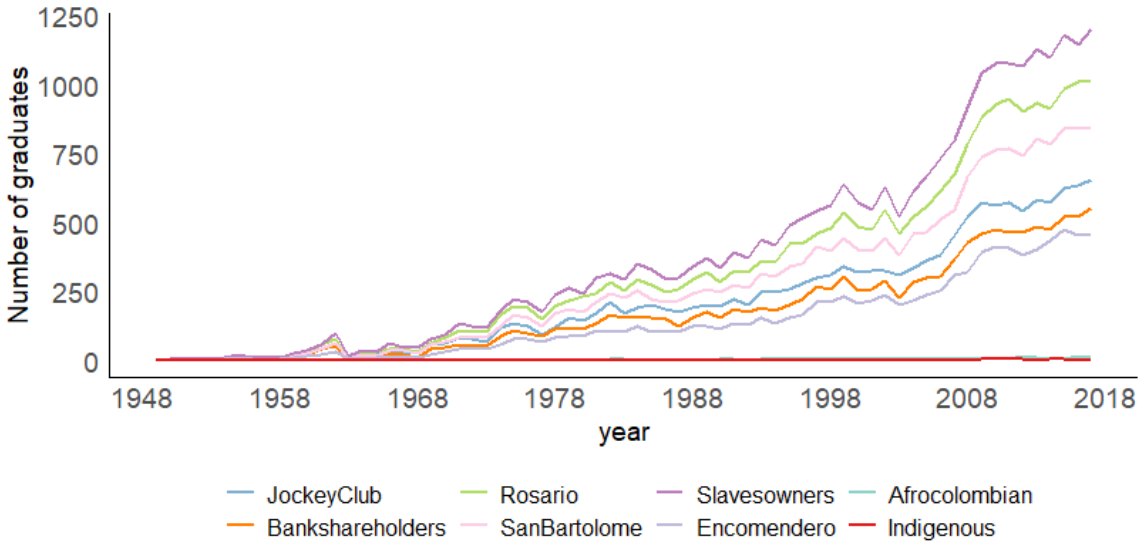
Table 6.2: Surnames and historical groups in SIMAT and Los Andes

Historical group	Num. of surnames	Num. rare surnames in SIMAT	Num. of graduates in Los Andes 1949-2018
Encomendero	363	65	607
slaves owners	1,607	615	3,799
Colegio Mayor del Rosario	711	255	2,895
Colegio Mayor San Bartolome	1,202	222	2,189
Bankshareholders	241	73	1,126
Jockey Club	309	107	1,661
Indigenous	131	113	47
Afro-Colombians	271	66	39

Note: The table provides the description of historical groups in Los Andes. Column 2 shows the number of surnames in each historical groups. Column 3 shows the number of rare surnames in SIMAT with the preferred definition of rare surname as those that have less than 2% of the number of holders of the most common surname in the sample (Rodriguez) and those with at least 3 holders to avoid possible misspellings errors. Column 4 shows the total number of graduates holding a rare surname in each historical group for the entire period. Sources: Historical groups from section 5.3. SIMAT and Graduates from Los Andes University.

Indeed, as shown in Fig. 6.2, most groups experienced considerable growth in the number of graduates in Los Andes since the University opened in 1948. The slave owners and the graduates from the colonial schools show the biggest increase, mainly driven by the higher number of rare surnames in these groups. However, the number of graduates with a surname that belongs to the Jockey Club, the encomenderos and the bankshareholders also increased over the years. On the contrary, the Afro-Colombians and Indigenous have a flat line over the period. As Table 6.2 reports, in the 73 years of Los Andes, only 47 graduates held an Indigenous surname and 39 graduates held an AfroColombian surname.

Figure 6.2: Number of graduates by groups, 1949-2018



Notes: The figure shows the total number of graduates by historical group in each year from 1949 to 2018. For a complete description of the historical groups see Section 5.3. Source: Graduates from Los Andes University.

The unequal patterns across the historical groups are also reflected in the year in which a surname of each historical group first appears in the graduates of Los Andes. As reported in Table 6.3, all elite groups have had at least one graduate in every year since the university started in 1949, while the ethnic groups had to wait more than 20 years in the case of the Afro-Colombians, and 36 years in the case of the Indigenous to appear among the graduates. Similarly, while the elite groups have access to a wide range of different degrees, having graduates in around 50 of the 97 programmes that Los Andes offers, the graduates from the ethnic groups are concentrated in less than 20 degrees. Noticeably, there is no difference among the most popular degree and across all historical groups the most popular programme is engineering.

Table 6.3: The degrees of the historical groups

Historical group	Year of first occurrence	Unique degrees	Most popular degree
Encomendero	1949	45	mechanical engineering
slaves owners	1949	65	industrial engineering
Colegio Mayor del Rosario	1949	62	industrial engineering
Colegio Mayor San Bartolome	1949	56	industrial engineering
Bankshareholders	1949	48	industrial engineering
Jockey Club	1949	50	industrial engineering
Indigenous	1976	15	chemical engineer
Afro-Colombians	1962	19	industrial engineering

Note: The table provides a description of historical groups in Los Andes. Column 2 shows the first year in which a surname that belongs to a historical group first appeared in the graduates of Los Andes. Column 3 shows the number of unique degrees found by historical group. Column 4 shows the most popular degree. Sources: Historical groups from section 5.3. SIMAT and Graduates from Los Andes University.

Similar to the findings of the previous chapter, these data confirm a pattern of segregation in access to Universidad de Los Andes in which the presence of elite groups and the absence of ethnic groups stand out. In the remainder of this chapter, I will focus on measuring social mobility in two generations of graduates.

6.4 Social mobility estimations

In general, what every study of social mobility aims at estimating is the correlation between a measure of the present social status (eg. income, education, wealth, health) and its values in the past, that is the correlation between adjacent generations. This assumes that the influence of the family on the social status of children can be measured by the correlation between the status of children and those of parents or grandparents (G. Becker, 1993).

Following the seminal work of G. Becker (1993), "the welfare of several consecutive generations of the same family would be closely linked whenever inheritability and investments are substantial" (p.212). This implies that in societies where inheritances play a major role, social mobility will be lower. In a similar way, in societies where the propensity to invest in future generations differs greatly among families, the influence of family background on the relative economic position of children is

greater. In general, at the cross-country level societies will differ in their levels of social mobility largely because of their social and political norms on inheritability, and at the country level, families will differ in their social mobility depending on the propensity to invest. Historically, social mobility could change then depending on social and political norms on inheritability and the propensity to invest of the families.

To measure the process of social mobility, most of the empirical literature assumes that social mobility follows an auto-regressive process and it is measured by estimating the coefficient β in a two-generation model (equation 6.1).[8]

$$Y_t = \alpha + \beta Y_{t-1} + \varepsilon_t \tag{6.1}$$

Where Y is a measure of social status, β is the persistence rate of the status, and 1-β is the social mobility rate. By definition $0<\beta<1$, where 0 represents a world of perfect social mobility in which the outcome of the offspring is not related to the outcome of the father, and 1 represents complete immobility because the status of the children can be perfectly predicted by the status of the father.[9] In the literature however, the estimations of the β cofficient range between 0.3 to 0.9.

The size of the β coefficient can also be interpreted as the influence of the family in the long run. G. Becker (1993) estimates that an intergenerational coefficient no larger than 0.4 implies that in no more than three generations all advantages and disadvantages of ancestors will disappear, suggesting a rapid regression to the mean, although empirical research has shown greater estimations of the β coefficient pointing towards slower rates of mobility (see for example Clark and Cummins (2014)

[8]Recent literature has also provided some evidence to consider an independent influence of grandparents on their grandsons not inherited through parents, see Long and Ferrie (2018) and for a discussion see Solon (2018).

[9]See also Black and Devereux (2011) for a discussion in the estimations of β.

and Ward (2021)).

The research on social mobility in Colombia has been focused on the correlation of income or education between two generations using contemporary survey data (Behrman et al., 2001; Angulo et al., 2012; Ramirez-Zuluaga, 2016). Hertz et al. (2007) find a fall in social immobility throughout 50 years. These results are similar to those from Behrman et al. (2001) and Angulo et al. (2012) that also find a slight change in the mobility patterns of the country in the past 30 years. Regarding the variables that these authors use to measure social status, Angulo et al. (2012) show that when social mobility is measured by the years of education, it can follow a different trajectory than when measured using a wealth index.[10] Overall, the literature concludes that it is more plausible to find regression towards the mean in education outcomes, than in income, and that it is difficult to move from one social class to another, especially at the extremes of the social ladder. As shown in Table 6.4, the estimated intergenerational coefficient is between 0.51 to 0.8, depending on the method and outcome. This chapter adds to the available estimates for Colombia but considers a long–term perspective, focusing on generations of graduates born between 1915 to 1949 and between 1950-1993.

Table 6.4: Intergenerational mobility estimations for Colombia

Article	Outcome	Intergenerational Mobility	Method	cohorts
Angulo et. al (2012)	education	0.74	intergenerational coefficient	1945 - 1985
	SES	0.6	intergenerational coefficient	1945 - 1985
Ramirez-Zuluaga (2016)	income	0.74	intergenerational coefficient	1945 - 1985
Dahan and Gaviria (1999)	education	0.59	sibling correlations	1977 - 1981
Hertz et. al (2007)	education	0.8	intergenerational coefficient	1928 - 1977
Behrman et. al (2001)	education	0.51	sibling correlations	1928 - 1977

Notes: The table summarises the results of the literature on social mobility in Colombia. It shows the results reported in the main tables of each paper for which an intergenerational coefficient is presented. A higher intergenerational mobility indicates lower social mobility.

[10]This indicates that there are some comparability limitations across studies. Additionally, this finding suggests that even when individuals invest more in education, it is difficult for them to move from one social class to another. This implies then that mechanisms such as education, which have typically been pointed out in the literature as an efficient policy to promote social mobility, in Colombia seem insufficient. This happens when the education system expands in coverage, but the same does not happen with quality, as in the case of Colombia (Camacho et al., 2017).

6.4.1 Families' status and Relative Representation

As mentioned above, the analysis of social mobility requires the definition of a historically comparable variable of social status between two or more generations. The ideal measure should show how social status transmits across generations, but occupation, permanent income or wealth are not always observed at the same time of life for the two generations studied. As a result, several studies use single observations of the father's occupation as a proxy which results in measurement error and attenuation in the intergenerational coefficient (Ward, 2021). However, unlike other countries such as the United States, in the case of Colombia, there is limited data that link parents to children and allow the observation of multiple father observations to better proxy for his permanent status (Abramitzky et al., 2020, 2021).

To measure the average social status of families or historical groups, I follow Clark and Cummins (2013) and assume that educational status is strongly linked to the underlying social status of a family. In this chapter, I measure educational status as graduating from an elite university. However, educational status is also affected by a random component i.e. some members of the family will not complete university or will graduate from a different university. The greater the variance of this random component, the lower the intergenerational coefficient of status for individual families. But when surnames belong to a specific social or ethnic group, the attenuation bias can be minimised, if the number of families within each social group is high as the average random component should tend to zero. Therefore, the intergenerational correlation of educational status will capture correctly the intergenerational correlation of social status.

However, it is likely that my current measure of educational status creates additional attenuation bias as it only includes information about Los Andes University. Los Andes is a sample of high quality-elite graduates, but there are several other

institutions in Bogota as well as in the rest of the country that are also elite universities. Unfortunately, obtaining data for other universities, including Universidad Nacional de Colombia (the most important public university in the country) has been difficult. Given that I do not observe the full set of elite universities, the estimations of this chapter should be considered a lower bound.

I compute the relative representation of each *rare* surname j (and the group of common surnames) in the graduates of Los Andes from 1949 to 2018, t, following equation 6.2. As discussed in the previous chapter, this ratio should be 1 for the common surnames, greater than 1 for high-status surnames and less than 1 for low-status surnames.[11] The relative representation at the group level is calculated as the average RR_j in a 5 year cohort by each group to reduce measurement error, as single-year observations may severely bias the estimations if there are transitory shocks or errors in the data (Ward, 2021).

$$RR_{j,t} = \frac{\text{Graduates with surname } j\text{, in } t/\text{Total graduates in } t}{\text{People in SIMAT with surname } j/\text{Total SIMAT}} \tag{6.2}$$

Fig. 6.3 shows the relative representation for each group from 1974 to 2018.[12] As expected, the RR for common surnames fluctuates around 1.0, which shows that these surnames have not been over, nor under-represented historically in Los Andes. The Indigenous surnames never get a RR greater than 1.0 during the 70 years, not even in contemporary data. Additionally, there are five cohorts where no graduate student from Los Andes has an Indigenous surname. The Afro-Colombians show a

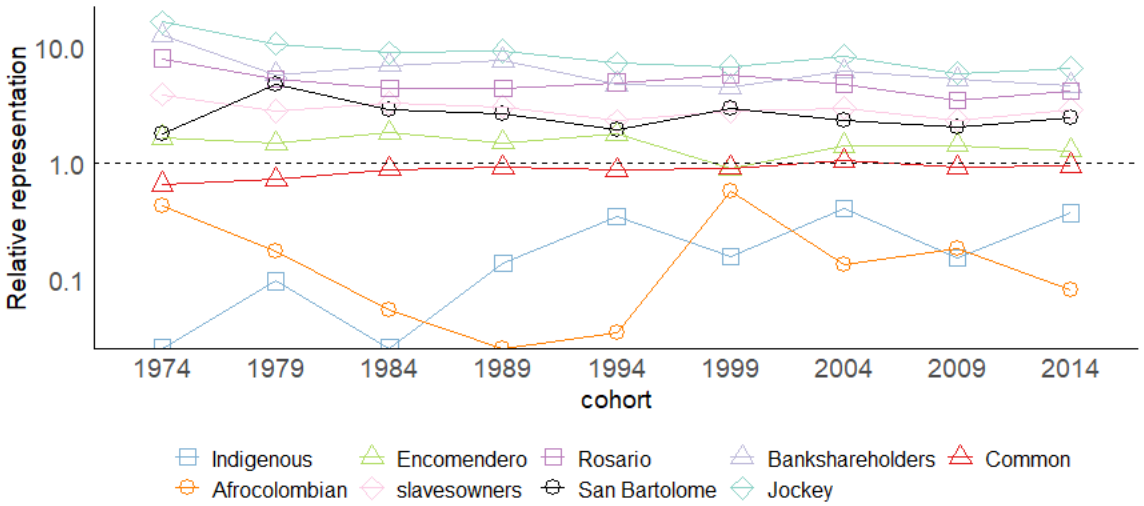
[11] I assume the denominator to be constant over time because I do not have information on the SIMAT for previous years. But in any case, the level of immigration in Colombia has been particularly low since the 19$^{\text{th}}$ century as the country did not witness the well-known big waves of migration as in Argentina, Uruguay, and Chile or Brazil.

[12] Fig. 6.6 shows the results for the full period. The RR shows considerable variance at the beginning of the university when groups like the bankshareholders and the Jockey club appear more than 30 times overrepresented among the graduates of the university.

similar pattern to the Indigenous, with 3 cohorts without any graduate from this group. But in 1962 there is one graduate in Los Andes with an AfroColombian surname that appears with 5 holders in the SIMAT, and therefore the estimated relative representation is very big and for that cohort, the average RR for the Afro-Colombians is higher than 15 (see Appendix Fig. 6.6). But apart from that exception, for the rest of the period, the Afro-Colombians always had an RR lower than one.

On the contrary, most of the elite groups, with the exception of the encomenderos, were always overrepresented in Los Andes by more than 2 times their share in SIMAT. It is noticeable, that surnames associated with Bankshareholders and the Jockey Club appear more than 5 times the average surname, during the 2014-2018 cohort, indicating that these groups have maintained their elite status for more than 100 years. Although the encomenderos appear with the smaller RR of all elite groups, they are still slightly overrepresented among the graduates of the University, which suggest that the group has also maintained its social status over time.

Figure 6.3: Average relative representation in 5-years cohort by group



Note: The figure reports the average relative representation at the group level where the relative representation is calculated as the average in a 5 years cohort by each of the historical groups and the 20 common surnames from 1974 to 2018. The dotted line indicates the relative representation of 1. The y-axis is log-scaled to ease interpretation. Sources: Historical groups from section 5.3. SIMAT and Graduates from Los Andes University.

These results are consistent with the literature that finds that lower classes

have slower social mobility and the greatest barriers to rising upward (Angulo et al., 2012). The elite groups in Los Andes show some convergence to the mean, while the ethnic groups show none. Finally, it is worth noticing that all the elite groups show a steady regression towards 1.0, but most of the elite groups remain systematically over-represented which suggests also very low social mobility for the elite, even for elite groups of the colonial period like the students from Rosario and San Bartolome.

Using surnames and their relative representation to measure social mobility has been criticised for averaging at surname or social group levels instead of providing a comparison between individuals (Chetty et al., 2014; Torche & Corvalan, 2018; Solon, 2018). Torche and Corvalan (2018) show that surname–level social mobility differs from individual-level social mobility because group-level averages capture both individual and group level characteristics that cannot be disentangled. Similarly, Chetty et al. (2014) discuss that group-level estimations can identify the behaviour of the group rather than across individuals. This distinction between individual and group persistence is especially important for the interpretation of the results and appropriate comparability with other studies.[13] In particular, individual results can differ from grouped results if there are different persistence patterns among defined groups of individuals, such as ethnic groups. As discussed in Chetty et al. (2014), marginalised racial groups in the US have different social mobility patterns in comparison with the average American. As social mobility is determined by a wide range of individual and socioeconomic characteristics, it can also be largely affected by certain group characteristics. For example, Ward (2021) recently showed that in the case of the US, the association between father and son between 1840-1910 increased from 0.3 to 0.7 after including Black families.

Beyond the discussion about the possibility to differentiate *within* from

[13]The literature is also aware of other comparability limitations of the study of social mobility. For example, it is still an open question the extent to which occupation-based measures can be compared to income-based measures or education (Pérez, 2019).

between group mobility, this study focus on the relative representation as a measure of relative social status due to the important data limitations of the country. With this measure I estimate the coefficient of intergenerational social mobility following equation 6.3 that estimates the correlation between two cohorts of graduates, dividing the complete sample into two cohorts of 35 years. This division allows me to compare two adjacent generations.

Given that the distribution of the relative representation is clustered around zero and it has a long right-hand tail, to produce a distribution closer to the normal I use a logarithmic transformation for the relative representation variable. I transform the variable by taking the log of $1 + RR_i$ to avoid $log(0) = undetermined$. Additionally, I am concerned with the outliers that are present in the distributions. In order to deal with this, I do all my estimations with the complete sample and truncate the sample.[14]

$$RR_{j,t} = \alpha + \beta * RR_{j,t-1} + \varepsilon_t \tag{6.3}$$

$RR_{j,t}$ is the logarithmic transformation of the relative representation of the surname j, in cohort t (1984-2018), and $RR_{j,t-1}$ is the relative representation of the surname j, in cohort $t-1$ (1949-1983). The β from these equations can be interpreted as the intergenerational elasticity between two generations. It is important to notice that as I am using surnames to follow families in the long run, this coefficient represents the *between group* coefficient of correlation. Additionally, I estimate this equation by each historical group. As previously discussed the small numbers of surnames in certain groups can limit the interpretation of the results. Similarly, as I focus only in Los Andes university, it is likely that the estimations of the coefficient

[14]I define an outlier as an observation that is 2 standard deviations above the mean relative representation of each social group. The results with the sample truncated are shown in the Table 6.7 and reveal that the results are not driven by outliers.

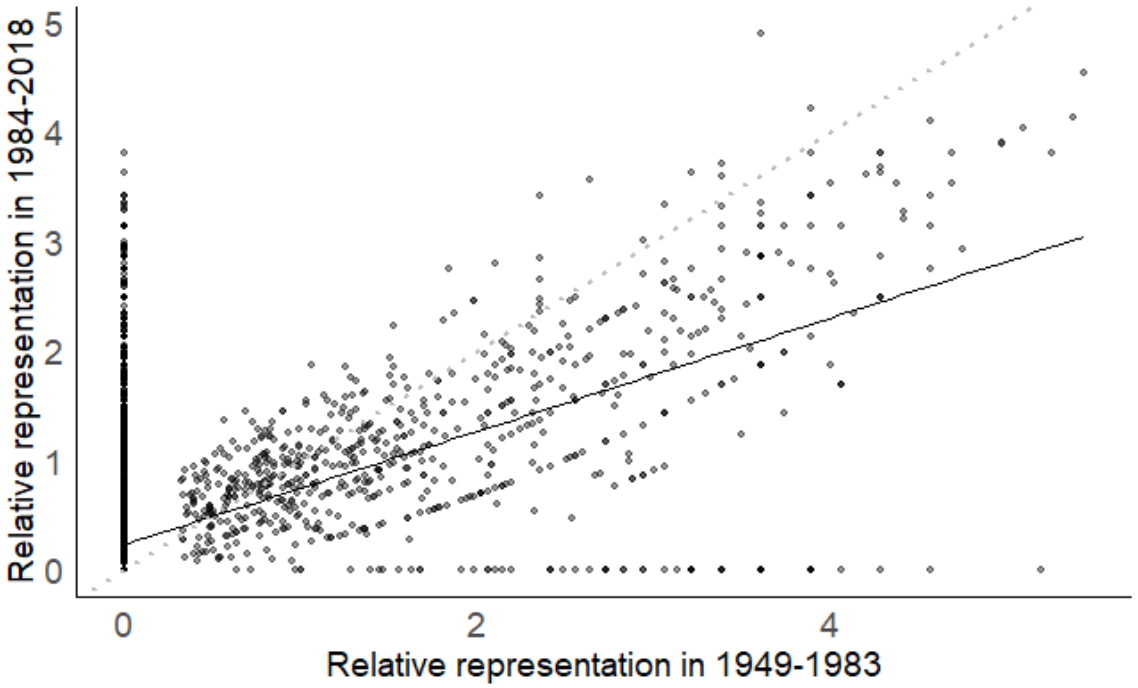
of intergenerational mobility are biased downwards.

6.5 Results

Fig. 6.4 shows the relationship between the relative representation in the generation of graduates in 1949-1983 and the relative representation in 1984-2018 for all rare surnames. The figure plots the average relative representation at the surname level, the predicted linear relationship and a hypothetical line of no mobility. In a completely immobile society the corresponding line would have a slope of 1. On the contrary, in a mobile society, the slope would approach zero. Intuitively, the 45-degree line implies that the RR of 1949-1983 perfectly predicts the RR in 1984-2018.

At the aggregated level the figure suggests that there is a strong and positive correlation between the relative representation in t and in $t-1$. In other words, families that were over-represented in $t-1$, are over-represented in t, suggesting low mobility in Los Andes. Nonetheless, when we compare the black line that corresponds to the predicted correlation with the grey dotted line that corresponds with perfect immobility, the figure shows that there is some degree of mobility between the two generations. It is also noticeable the number of zeros that appear in both generations, which is in line with the exclusivity of the University. The zeros are particularly important in the first generation when several families do not appear despite having a high relative representation in the second period. As there are fewer cases of families appearing in the first cohort and disappearing in the second, the presence of zeros therefore could drive downwards the estimation of the coefficient of intergenerational mobility.

Figure 6.4: Association between two generations



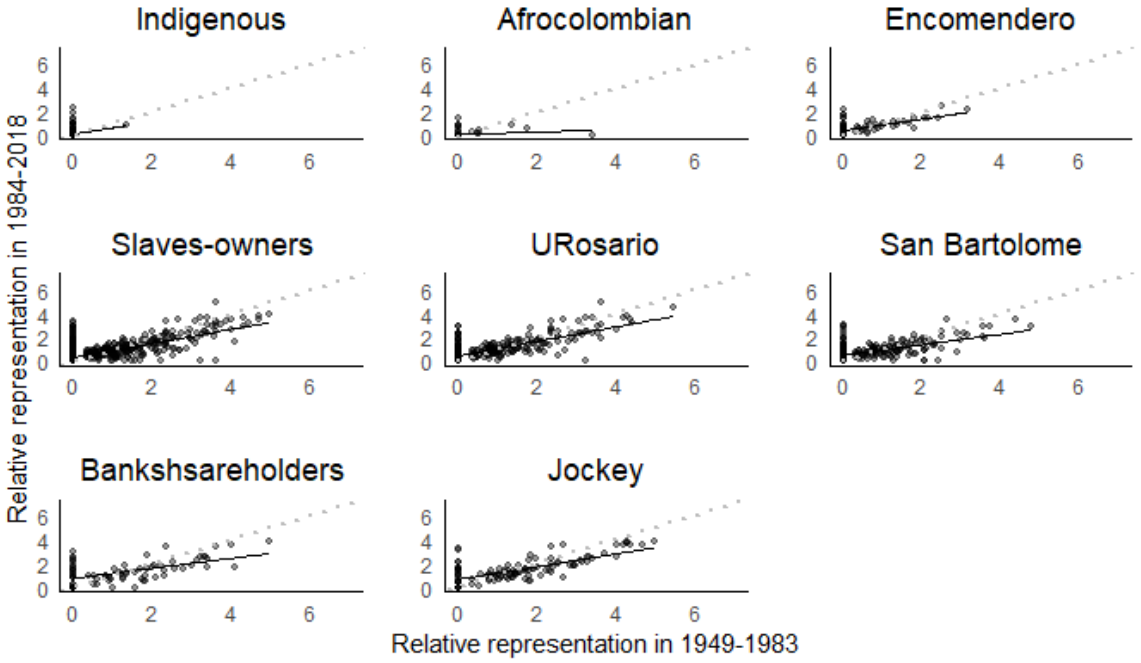
Note: The graph shows the intergenerational correlation between graduates of Los Andes in the 1949-1983 generation and the 1984-2018 generation. Each dot represents the average relative representation of a surname in each generation following Eq. (6.2) and the relative representation corresponds to the Log(1+RR_g). The grey dotted line represents a situation of perfect immobility while the black line is the predicted linear correlation in the data. Sources: Graduates from Los Andes University.

Fig. 6.5 shows the relationship between the relative representation in the generation of graduates in 1949-1983 and the relative representation in 1984-2018 by historical groups. Similar to the findings in Chapter 5, the results indicate that there are different patterns of social mobility by historical groups. The Indigenous show low mobility mainly because the family *Tenjo*, the only family that shows up in the 2 generations, appears with a RR higher than 1 in both periods. The rest of the Indigenous surnames that appear in Los Andes belong to graduates in the period 1984-2018. Nonetheless, the total number of Indigenous surnames in the whole period of study is very low. The Afro-Colombians show a distinct pattern. The predicted correlation of this group suggests almost perfect mobility but again this result is driven by the low number of AfroColombian surnames present in Los Andes. Additionally, the results are also explained by the family *Cuca*, which appears with

a very high relative representation in the 1949-1983 period, but disappears in the second generation.

Most of the elite groups show low social mobility, in particular the Slavesowners, the graduates from Rosario and the members of the Jockey Club. These elite groups have a predicted linear correlation that is close to a situation of perfect immobility. However, all elite groups have a considerable number of families that only appear in the second generation which pushes the correlation downwards. The encomenderos and the bankshareholders are the elite groups that appear with the highest social mobility in the sample. The families that belong to the bankshareholders seem to have higher relative representation in the first generation in comparison to the second generation, which explains the almost horizontal shape of the predicted line. This is also shown in Fig. 6.6, where the bankshareholders have a RR of 45 in the cohort of 1949.

Figure 6.5: Intergenerational elasticity in Los Andes University



Note: The graph shows intergenerational correlation between graduates of Los Andes in the 1949-1983 generation and the 1984-2018 generation by each historical group. Each dot represents the average relative representation of a surname in each generation following Eq. (6.2) and the relative representation corresponds to the Log(1+RR_g). The grey dotted line represents a situation of perfect immobility while the black line is the predicted linear correlation in the data. Sources: Graduates from Los Andes University.

Table 6.5 shows the result from the estimation of the equation 6.3 with the complete sample as well as for each of the historical groups. The first column suggests that, overall, there is a strong, positive, and significant intergenerational elasticity between the two cohorts (0.52). This intergenerational elasticity is lower in comparison to other estimations for Colombia (see Table 6.4), but implies that it would take more than 4 generations for a family to regress to the mean status. Columns 2 to 9 report the results for each group. The differences in the coefficients (and the R-square) support the idea of different social mobility patterns at the group level and allow us to give another look at the comparative social mobility between these historical groups.

The results show a significant high and positive intergenerational elasticity of social status for almost all the groups, with the exception of the Afro-Colombians who have a small and non-significant intergenerational elasticity explained by the small number of observations. The surnames of the Jockey club, the slaves owners and Rosario behave very similarly and have a coefficient higher than 0.5, suggesting that these groups have the lowest social mobility. In the case of the Colegio Mayor San Bartolome students, founders of banks and encomenderos we observe a smaller coefficient (around 0.45) indicating more mobility in comparison to the whole sample and the other elite groups. Although these results suggest that the pre-modern elite is more mobile than the modern elite, it is worth highlighting that while the relative representation has declined over time for the modern groups as shown in Fig. 6.3, the relative representation has remained stable (although lower) in the pre-industrial groups. The results are robust and not driven by the different definitions of rarity nor by outliers as shown in Appendix Table 6.7 where I repeat the estimation of the equation truncating the Relative Representation.

Table 6.5: Intergenerational elasticity in Los Andes graduates.

	log (1+RR 1984-2018)								
	All	Indigenous	AfrCol	Encomenderos	slaves owners	Rosario	San Bartolo	Banksharehold	Jockey
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A	Rarity: 3-150 holders								
log (1+RR 1949-1983)	0.512***	0.504	0.093	0.521***	0.594***	0.606***	0.436***	0.391***	0.504***
	(0.011)	(0.311)	(0.071)	(0.113)	(0.028)	(0.042)	(0.054)	(0.086)	(0.056)
Constant	0.229***	0.152***	0.089**	0.427***	0.358***	0.527***	0.563***	0.938***	0.825***
	(0.008)	(0.040)	(0.038)	(0.117)	(0.037)	(0.071)	(0.074)	(0.179)	(0.124)
Observations	6,722	105	57	36	471	178	153	50	83
Adjusted R^2	0.249	0.015	0.012	0.368	0.489	0.537	0.295	0.288	0.494
Panel B	Rarity: 3-372 holders								
log (1+RR 1949-1983)	0.517***	0.510*	0.102	0.485***	0.588***	0.606***	0.441***	0.419***	0.535***
	(0.010)	(0.010)	(0.301)	(0.079)	(0.024)	(0.035)	(0.044)	(0.071)	(0.050)
Constant	0.232***	0.144***	0.096***	0.411***	0.347***	0.483***	0.526***	0.778***	0.683***
	(0.007)	(0.007)	(0.037)	(0.075)	(0.029)	(0.053)	(0.054)	(0.130)	(0.103)
Observations	7,125	113	66	65	615	255	222	73	107
Adjusted R^2	0.256	0.016	0.020	0.363	0.489	0.536	0.314	0.319	0.513
Panel C	Rarity: 3-500 holders								
log (1+RR 1949-1983)	0.519***	0.510*	0.190***	0.500***	0.590***	0.611***	0.447***	0.440***	0.545***
	(0.010)	(0.301)	(0.067)	(0.070)	(0.023)	(0.033)	(0.041)	(0.063)	(0.047)
Constant	0.234***	0.144***	0.102***	0.406***	0.342***	0.461***	0.520***	0.708***	0.643***
	(0.007)	(0.037)	(0.036)	(0.063)	(0.027)	(0.047)	(0.049)	(0.106)	(0.091)
Observations	7,244	113	70	80	688	291	251	91	121
Adjusted R^2	0.259	0.016	0.093	0.385	0.493	0.544	0.323	0.349	0.530

Notes: Estimation are based on an OLS model. The results show the intergenerational elasticity between the 1949-1983 and 1984-2018 cohorts. The first column shows the aggregated the results using all rare surnames in Los Andes. Columns 2 to 9 report the results for each individual group. Panel A shows the estimations based on rare surnames defined as those that are held by 3 to 150 individuals in SIMAT. Panel B reports the results of my preferred definition of rarity and discussed in Section 6.4. Panel C shows the results based on rare surnames defined as those held by 3-500 holders in SIMAT. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1 indicate statistical significance.

Finally, to disentangle the differences in social mobility among elite groups I estimate the coefficient of intergenerational elasticity for those elite families that appear both in pre-modern and modern elite groups and those that appear in only one historical period. Similar to the results in Chapter 5, these findings show that long-surviving elites have lower social mobility and higher persistent rates in comparison to families that were unable to enrol in the modern elite groups or elite families that appear only in the modern period. Long-surviving families have an intergenerational coefficient close to 0.55, while families that only belong to the modern period have a coefficient of 0.33. However, it is important to highlight that there are less than 40 families only in the modern period which probably biases downwards the coefficient, as in the case of the Afro-Colombians.

Table 6.6: The long-lasting elite: Intergenerational elasticity

	log (1+RR 1984-2018)		
	Multiperiod elite (1)	Only pre-industrial elite (2)	Only modern elite (3)
Panel A	Rarity: 3-150 holders		
log (1+RR 1949-1983)	0.518***	0.545***	0.300**
	(0.057)	(0.029)	(0.131)
Constant	0.696***	0.388***	1.248***
	(0.130)	(0.032)	(0.215)
Observations	82	586	28
Adjusted R^2	0.499	0.382	0.135
Panel B	**Rarity: 3-372 holders**		
log (1+RR 1949-1983)	0.543***	0.539***	0.325**
	(0.047)	(0.025)	(0.126)
Constant	0.583***	0.380***	1.091***
	(0.095)	(0.026)	(0.192)
Observations	115	758	34
Adjusted R^2	0.533	0.380	0.146
Panel C	Rarity: 3-500 holders		
log (1+RR 1949-1983)	0.552***	0.540***	0.337***
	(0.043)	(0.024)	(0.123)
Constant	0.551***	0.373***	1.019***
	(0.080)	(0.024)	(0.183)
Observations	136	824	37
Adjusted R^2	0.553	0.383	0.152

Notes: Estimation are based on an OLS model. The results show the intergenerational elasticity between the 1949-1983 and 1984-2018 cohorts. The first column shows the results for elite surnames that appear in more than one historical period, Column 2 reports the results for surnames that appear only in the pre-industrial elite and Column 3 shows the results for families that appear only in the modern period. Panel A shows the estimations based on rare surnames defined as those that are held by 3 to 150 individuals in SIMAT. Panel B reports the results of my preferred definition of rarity and discussed in Section 6.4. Panel C shows the results based on rare surnames defined as those held by 3-500 holders in SIMAT. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1 indicate statistical significance.

In general the results are consistent with what other social mobility research finds for Colombia, with an overall intergenerational coefficient of 0.52 (Angulo et al., 2012; Behrman et al., 2001; Ramirez-Zuluaga, 2016; Hertz et al., 2007). Additionally, the findings support the idea that there are different mobility patterns by groups in which social mobility is very slow in the upper and lower parts of the social status ladder. All the elite groups seem to experience slow social mobility while for the ethnic

groups the results are not significant given their scarce participation in Los Andes. However, the very low participation of ethnic groups in an elite institution indicates that there is a strong influence of family background on the relative economic position of future generations.

6.6 Conclusions

According to contemporary measures, Colombia is one of the least mobile countries in the world. However, it has been difficult to track social mobility in the long run. This research has filled this gap using surnames and their relative representation to link the past to the present. Following rare surnames from Indigenous groups, 17$^{\text{th}}$ century Spaniard colonial officers (*encomenderos*), African enslaved people, 17$^{\text{th}}$ and late 18$^{\text{th}}$ century members of privileged families with access to higher education, 19$^{\text{th}}$ century slaves owners, and members of different social and business elites of the late 19$^{\text{th}}$ and the beginning of the 20$^{\text{th}}$ centuries, I observe generational changes in the relative representation of these historical social groups among the alumni of Los Andes University from 1949 to 2018.

I estimate the inter–generational correlation coefficient between the relative representation of 2 cohorts from 1949 to 2018, at the surname and the group–level. I find that families that were over-represented in Los Andes in the past, are still over-represented, indicating low social mobility patterns. Overall I find an intergenerational coefficient of 0.52 for the whole sample. Using historical groups I link the past to the present and find that low mobility is especially noticeable for the elite families that appear in elite groups of the colonial period as well as in the late 19$^{\text{th}}$ century. I estimate intergenerational coefficients that range from 0.45 to 0.6 for the different elite groups. Additionally, I find that long-surviving elite families show lower social mobility in comparison to families that only appear in the pre-modern period but were

unable to enrol in other elite groups in the late 19th century. Ethnic surnames associated with Indigenous and Afro-Colombians however show different behaviour. These groups have a non-significant intergenerational coefficient, explained mainly by their persistent underrepresentation during the whole period since Los Andes University opened.

Although a comparison with other countries is limited by my measure of social status, my estimations suggest that intergenerational mobility was lower in Los Andes than in the US in the 1930s (Feigenbaum, 2018; Olivetti & Paserman, 2015). My estimations are lower than those for the educational system in Sweden and England estimated by Clark and Cummins (2013, 2014) but this could be due to the attenuation bias caused by only using information from Los Andes.[15]

In comparison to other studies of social mobility in Colombia, my estimates are generally lower but my results support the finding that in Colombia, social mobility is very slow in the upper part of the social status ladder. Even if elite groups are slowly regressing to the mean, the patterns of the older elites like the *encomenderos* indicate that contrary to Becker's hypothesis, all the historical groups studied will take more than 3 generations to achieve convergence to the mean. These findings also suggest that throughout history, elite groups have conserved their high status by maintaining exclusive access to high-quality education, as in the case of the students from Colegio Mayor del Rosario and San Bartolome. These historical elite groups have been always over-represented in elite educational institutions, while Afro-Colombians and Indigenous have rarely accessed these institutions.

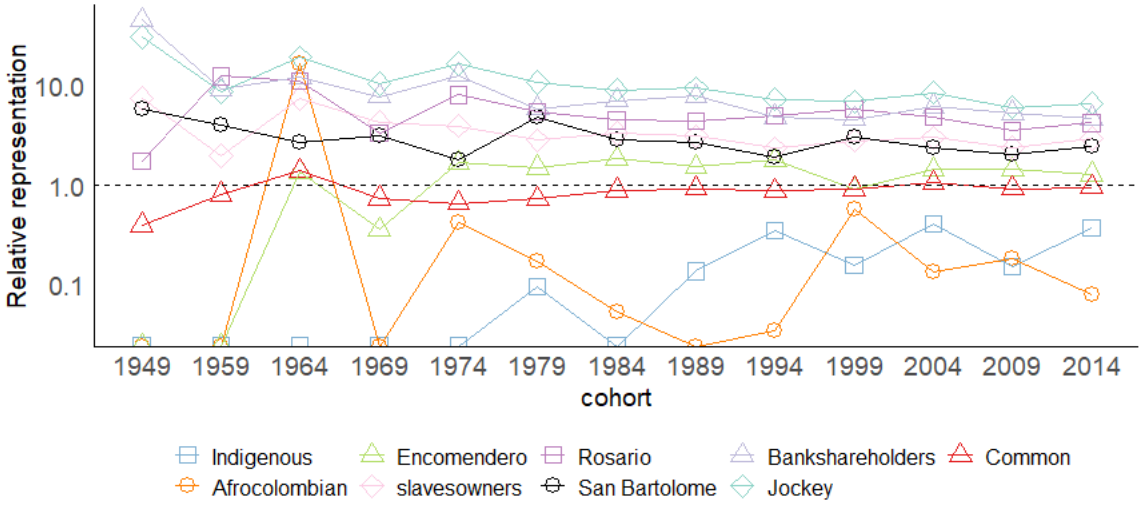
[15]Using educational information Clark and Cummins (2013, 2014) estimate an interegenerational mobility coefficient of 0.8 for England during 1950-2012 and of 0.66 for Sweden during the same period.

6.A Appendix

6.A.1 Complete period: 1949-2018

Fig. 6.6 shows the results for the period 1949 t0 2018. The *RR* shows considerable variance at the beginning of the university when groups like the bankshareholders and the Jockey club appear more than 30 times overrepresented in the graduates of the university.

Figure 6.6: Average relative representation by group.



Note: The figure reports the average relative representation at the group level where the relative representation is calculated as the average in a 5 years cohort by each of historical groups and the 20 common surnames from 1949 to 2018. The dotted line indicates the relative representation of 1. The y-axis is log-scaled to ease interpretation. Sources: Historical groups from section 5.3. SIMAT and Graduates from Los Andes University.

6.A.2 Robustness: Different thresholds of rarity

To address potential selection in our rare surnames I consider other definitions of rarity. Similarly, given the long right-tails of the relative representation I present estimations truncated at 2 standard deviations of the mean. The results reveal that the patterns remain robust.

Table 6.7: Intergenerational elasticity in Los Andes graduates - truncated OLS

	log (1+RR 1984-2018)								
	All	Indigenous	AfrCol	Encomenderos	slaves owners	Rosario	San Bartolo	Banksharehold	Jockey
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A	Rarity: 3-150 holders								
log (1+RR 1949-1983)	0.497***	0.504	0.112	0.522***	0.540***	0.517***	0.409***	0.321***	0.396***
	(0.012)	(0.311)	(0.077)	(0.114)	(0.028)	(0.042)	(0.054)	(0.086)	(0.054)
Constant	0.224***	0.152***	0.088**	0.427***	0.359***	0.553***	0.560***	0.961***	0.890***
	(0.007)	(0.040)	(0.038)	(0.117)	(0.033)	(0.063)	(0.068)	(0.160)	(0.103)
Observations	6,722	105	57	36	471	178	153	50	83
Adjusted R^2	0.217	0.015	0.020	0.364	0.444	0.464	0.267	0.209	0.393
Panel B	Rarity: 3-372 holders								
log (1+RR 1949-1983)	0.683***	0.474	0.164*	0.566***	0.719***	0.731***	0.527***	0.460***	0.632***
	(0.015)	(0.320)	(0.083)	(0.107)	(0.035)	(0.054)	(0.062)	(0.119)	(0.089)
Constant	0.840***	0.785***	0.746***	0.934***	0.898***	0.985***	1.011***	1.241***	1.148***
	(0.005)	(0.025)	(0.021)	(0.057)	(0.020)	(0.039)	(0.040)	(0.101)	(0.082)
Observations	7,125	113	66	65	615	255	222	73	107
Adjusted R^2	0.221	0.010	0.042	0.299	0.410	0.416	0.244	0.163	0.319
Panel C	Rarity: 3-500 holders								
log (1+RR 1949-1983)	0.506***	0.510*	0.223***	0.500***	0.539***	0.526***	0.424***	0.376***	0.444***
	(0.011)	(0.301)	(0.071)	(0.071)	(0.023)	(0.032)	(0.041)	(0.063)	(0.046)
Constant	0.228***	0.144***	0.098***	0.406***	0.350***	0.497***	0.523***	0.744***	0.715***
	(0.007)	(0.037)	(0.036)	(0.063)	(0.024)	(0.042)	(0.046)	(0.097)	(0.078)
Observations	7,244	113	70	80	688	291	251	91	121
Adjusted R^2	0.230	0.016	0.113	0.381	0.452	0.474	0.302	0.275	0.434

Notes: Estimation are based on an OLS model with truncated values of the relative representation. The RR is truncated to 2 standard deviations above the mean. The results show the intergenerational elasticity between the 1949-1983 and 1984-2018 cohorts. The first column shows the aggregated the results using all rare surnames in Los Andes. Columns 2 to 9 report the results for each individual group. Panel A shows the estimations based on rare surnames defined as those that are held by 3 to 150 individuals in SIMAT. Panel B reports the results of my preferred definition of rarity and discussed in Section 6.4. Panel C shows the results based on rare surnames defined as those held by 3-500 holders in SIMAT. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1 indicate statistical significance.

Table 6.8: The long-lasting elite: Intergenerational elasticity - truncated OLS

	log (1+RR 1984-2018)		
	Multiperiod elite (1)	Only pre-industrial elite (2)	Only modern elite (3)
Panel A	Rarity: 3-150 holders		
log (1+RR 1949-1983)	0.418***	0.492***	0.254*
	(0.057)	(0.029)	(0.126)
Constant	0.771***	0.390***	1.208***
	(0.111)	(0.030)	(0.187)
Observations	82	586	28
Adjusted R^2	0.396	0.330	0.102
Panel B	Rarity: 3-372 holders		
log (1+RR 1949-1983)	0.670***	0.628***	0.331*
	(0.085)	(0.035)	(0.191)
Constant	1.060***	0.925***	1.440***
	(0.078)	(0.018)	(0.137)
Observations	115	758	34
Adjusted R^2	0.350	0.298	0.058
Panel C	Rarity: 3-500 holders		
log (1+RR 1949-1983)	0.552***	0.540***	0.337***
	(0.043)	(0.024)	(0.123)
Constant	0.551***	0.373***	1.019***
	(0.080)	(0.024)	(0.183)
Observations	136	824	37
Adjusted R^2	0.553	0.383	0.152

Notes: Estimations are based on an OLS model with truncated values of the relative representation. The *RR* is truncated to 2 standard deviations above the mean. The results show the intergenerational elasticity between the 1949-1983 and 1984-2018 cohorts. The first column shows the results for elite surnames that appear in more than one historical period, Column 2 reports the results for surnames that appear only in the pre-industrial elite and Column 3 shows the results for families that appear only in the modern period. Panel A shows the estimations based on rare surnames defined as those that are held by 3 to 150 individuals in SIMAT. Panel B reports the results of my preferred definition of rarity and discussed in Section 6.4. Panel C shows the results based on rare surnames defined as those held by 3-500 holders in SIMAT. Standard errors in parentheses and *** p<0.01, ** p<0.05, * p<0.1 indicate statistical significance.

7

Conclusions

This thesis contributes to our understanding of fertility decline and the persistence of social status in twentieth-century Colombia, two topics about which there is still so much to learn. In this section, I reexamine the questions and findings of this thesis. I highlight some broader lessons of the results and, in some parts, I stop to address their policy implications. Additionally, I discuss the limitations of the scope of this thesis, and I try to shed light on new avenues of research.

The first part of the thesis addressed the fertility decline. It presented a new analysis of fertility before and during the fertility decline at the sub-national

level. The findings of Chapter 2 revealed significant variations in fertility before the fertility decline. Although I did not explore the reasons behind this difference, I speculate that municipalities that had higher fertility were closer to the frontier that was expanding across Colombia since the late 19th century. This suggests that fertility norms differed as a consequence of different historical legacies. Supporting the findings of the European Fertility Project, this novel perspective of the Colombian case reveals that fertility declined simultaneously across different areas, despite well-known regional economic disparities. The main exception was the indigenous communities that did not undergo a fertility decline during this period.

The simultaneous decline across regions suggests that sudden technological or cultural shocks might explain the decline. Therefore, Chapter 3 looked into the potential effects of family planning. As Miller (2010) suggested, knowledge of and access to Profamilia had a limited role in the fertility decline of the country. Although the importance of the promotion of contraception through radio has been ruled out in this thesis, this should not be interpreted as if media and communication technologies cannot affect fertility decisions. On the contrary, the results suggest that the potential effects of media on fertility have to do more with the promotion of new social norms and opportunities than with the diffusion of information about family planning. In the 1960s Colombia was more connected than ever before and the radio likely played a key role in the nation-building process. So it is reasonable to think that the concerns of the government regarding population growth were diffused to families through radio, TV and newspapers even before a population policy was formally established.

Although a definitive cause behind the rapid fertility decline of Colombia is still missing, in Chapter 4 I studied in detail the relationship between maternal education and fertility. The chapter provided new stylised facts on the national, sub-national and individual-level. Nationally, fertility decline cannot be explained

through the direct effects of the increases in educational coverage because fertility also declined smoothly for non-educated women. This finding is relevant as peer effects were ruled out. These results make sense in a country like Colombia, as it has a strong segregation in the educational system. The little interaction across social classes in education moderates the potential peer effects. However, it is left to future research to investigate what drove the fertility decline in uneducated women. As for individual determinants, the results show that access to secondary or higher education resulted in overall lower fertility and a higher probability of remaining childless.

Part two of the thesis presented novel evidence of unequal opportunity to access education and how social status has persisted across generations. The findings in Chapter 5 suggest that educational segregation reproduces patterns of social exclusion that are rooted in Colonial institutions that benefited elite groups. For example, while colonial requirements to enter higher education were relaxed after 1840, the results indicate that *de jure* segregation in access to higher education continues today, and elite groups and ethnic groups rarely mix. This segregation is reflected in the persistent under-representation of historically excluded social groups (Afro-Colombians and indigenous) in high–quality education. Also, contemporary segregation seems to be related to the high levels of homogamy in indigenous people and Afro-Colombians.

Chapter 6 focused on a case study of historical social mobility in access to high-quality education. The findings indicate low social mobility in Los Andes University and corroborate the exclusion of ethnic groups. How much more or less mobile was Los Andes University in comparison to other elite universities is difficult to estimate. It is likely, however, that my estimations represent a lower bound, as I do not observe all high-quality universities. Also, the results may differ when looking at other high-quality but public universities such as Universidad Nacional or Universidad de Antioquia. The relatively low cost of these universities and their good reputation

(*goodwill*) can promote more integration across different social groups.

Chapter 2, Chapter 5 and Chapter 6 contribute to our understanding of indigenous communities in Colombia and the scarce literature on the social change of indigenous populations. Indigenous communities are mostly located in isolated areas where the effects of economic growth, modernisation and education are likely to be different from the rest of the country. Although far from comprehensive, these chapters provide an initial empirical view of their fertility during the fertility decline and of their presence in the contemporary educational system. More research is required on the fertility of indigenous women and other social outcomes, not only in Colombia but in the rest of the Americas. I believe this will be fertile ground for future research.

Some questions remained unanswered. Fertility patterns before 1958 are still largely unknown at the municipal level. Similarly, we do not have consistent series of vital statistics in Colombia and, in particular, we do not have consistent information on the number of births by the department for the whole 20^{th} century. This type of information is crucial to understanding how fertility evolved before the fertility transition, while urbanisation and industrialisation expanded. I hope that this thesis motivates other researchers to search and collect censuses at the individual level for years before 1964 and to estimate fertility rates.

Also, this thesis does not look into migration in detail due to data constraints. Migration was a key factor during the modernisation process of the country, especially after the 1870s, but there were important internal migration waves in the 1920s and 1950s. In most of the thesis, the analyses of fertility and social mobility are made using the place of residence instead of the place of birth. Information about the place of birth is available in the censuses but we do not observe when they migrate. Looking into the relationship between fertility and migration can elucidate

how fertility patterns diffuse –or not– from one place to another.

For the social mobility study, we only have the place in which students are enrolled. Therefore, the analysis of segregation in the school system and its influences on social mobility has been made at the national level. As extensively discussed in Chapter 2 the regional approach in a country like Colombia deserves further exploration. Likely, both access to high-quality schooling and mobility patterns vary widely across the regions.

The results of this thesis issue two warnings. First, the tradeoff between fertility and education was stronger for young women than for older women in 1973 and recent research suggests that this seems to have changed little (Batyra, 2016; Urdinola & Ospino, 2015). In particular, after 1990 teenage fertility started increasing across Latin America and this increase has been more prevalent among teenagers with low levels of education. In a country with low social mobility and limited opportunities for young women, women may value investing in their education only if they think they will be able to succeed in society. Otherwise, young women can regard motherhood as a source of personal realisation. The results of this thesis suggest that the government must prioritise the promotion of new social roles for girls and teenagers while ensuring access to secondary education as a way to reduce teenage pregnancy in Colombia.

Second, Colombia, as a multi-ethnic country, should promote educational institutions in which ethnic groups and different social classes can meet. Historically, these institutions have been scarce and continue to be so. The results indicate that today, public schools show more association between ethnic groups and the historical elite in comparison to private institutions. To reduce the barriers to social and economic mobility, it is crucial to reduce the segregation of the educational system and to direct more resources towards affordable and better quality education targeted

specifically to historically marginalised groups.

Finally and looking into the future, I hope that this thesis lays the foundations of an all-Latin American project. This is crucial to understand the historical socioeconomic differences among (and within) Latin American countries and to this end we can make use of the digitisation of historical censuses and the availability of genealogical records to study these socioeconomic transformations in a long-term perspective. As part of this initiative, I started the "Latin American Fertility Project" which shows how to calculate age-specific and total fertility rates to study the fertility transition in several countries using the available data from IPUMS and Human Life-Table Database.[1] The results suggest that in most countries fertility transition came before policy interventions. With the census data it is possible to expand these estimations to calculate the TFR by area of residence, level of schooling or the occupational status of the mother, and to compare fertility changes across countries or even borders. The still unexploited genealogical data also provides a source of abundant information that can be used to measure long-run social mobility and to compare Latin American countries systematically.

[1] I released the implementation of the OCM to calculate Total Fertility Rates (TFR) by country over 14 years under an open source license (`https://github.com/jje90/ASFR_TFR`).

References

Aaronson, D., Lange, F. & Mazumder, B. (2014). Fertility transitions along the extensive and intensive margins. *American Economic Review*, *104*(11), 3701–24.

Abramitzky, R., Boustan, L., Eriksson, K., Feigenbaum, J. & Pérez, S. (2021). Automated linking of historical data. *Journal of Economic Literature*, *59*(3), 865–918.

Abramitzky, R., Mill, R. & Pérez, S. (2020). Linking individuals across historical sources: A fully automated approach. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *53*(2), 94–111.

Acemoglu, D., Bautista, M. A., Querubin, P. & Robinson, J. (2007). *Economic and Political Inequality in Development: The Case of Cundinamarca, Colombia* (Working Paper No. 13208). NBER: National Bureau of Economic Research.

Acemoglu, D., García-Jimeno, C. & Robinson, J. A. (2012). Finding El Dorado: slavery and long-run development in Colombia. *Journal of Comparative Economics*, *40*(4), 534–564.

Acemoglu, D., Johnson, S. & Robinson, J. A. (2005). Institutions as a fundamental cause of long-run growth. In P. Aghion & S. Durlauf (Eds.), *Handbook of economic growth* (Vol. 1, pp. 385–472).

Acosta-Ordoñez, K. (2014). *Estimación indirecta de la tasa de mortalidad infantil en Colombia, 1964-2008* (Working Paper No. 199). Documentos de Trabajo Sobre Economía Regional y Urbana: Banco de la República de Colombia.

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V. & Zhuravskaya, E. (2015). Radio and the Rise of the Nazis in Prewar Germany. *The Quarterly Journal of Economics*, *130*(4), 1885–1939.

Ager, P., Boustan, L. & Eriksson, K. (2021). The intergenerational effects of a large wealth shock: white southerners after the civil war. *American Economic*

Review, *111*(11), 3767–94.

A'Hearn, B., Baten, J. & Crayen, D. (2009). Quantifying quantitative literacy: Age heaping and the history of human capital. *The Journal of Economic History*, *69*(3), 783–808.

Alvaredo, F. & Londoño Velez, J. (2013). *High incomes and personal taxation in a developing economy: Colombia* (Working Paper No. 12). CEQ: Commitment to Equity.

Alvarez, D. (2018). *Explaining informality: Extractive states and the persistent incentives for being lawless.* (Mimeo)

Alzúa, M. L. & Velázquez, C. (2017). The effect of education on teenage fertility: causal evidence for Argentina. *IZA Journal of Development and Migration*, *7*(1).

Angrist, J. D. & Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review*, *88*(3), 450–477.

Angulo, R., Azevedo, J. P., Gaviria, A. & Páez, G. N. (2012). *Movilidad social en Colombia* (Working Paper No. 30). Documento CEDE: Universidad de Los Andes.

Anselin, L. (1995). Local indicators of spatial association—lisa. *Geographical analysis*, *27*(2), 93–115.

Arias-Valencia, M. M. (2001). Comportamiento reproductivo en las etnias de Antioquia-Colombia. *Salud Pública de México*, *43*(4), 268–278.

Arias-Valencia, M. M. (2005). Determinantes proximos de la fecundidad: comportamiento reproductivo de las indigenas Chamibida de Antioquia, Colombia. *Cad. Salud Publica*, *21*(4), 1087–1098.

Arrazola, R. (1970). *Palenque: primer pueblo libre de América.* Cartagena de Indias: Ediciones Hernández.

Avery, C., St. Clair, T., Levin, M. & Hill, K. (2013). The 'own children' fertility estimation procedure: A reappraisal. *Population Studies*, *67*(2), 171–183.

Axinn, W. G. & Barber, J. S. (2001). Mass Education and Fertility Transition. *American Sociological Review*, *66*(4), 481–505.

Babiarz, K. S., Ma, P., Miller, G. & Song, S. (2020). *The Limits and consequences of population policy: Evidence from China's Wan Xi Shao campaign* (Working Paper No. 25130). NBER: National Bureau of Economic Research.

Baez, J. (2008). *Does more mean better? sibling sex composition and the link between family size and children's quality* (Working Paper No. 3472). Institute of Labor Economics: IZA.

Bailey, J. (1973). An evaluative look at a family planning radio campaign in Latin America. *Studies in Family Planning*, *4*(10), 275–278.

Bailey, J. & Cabrera, E. (1981). Campañas de radio y planificacion de la familia en Colombia, 1971-1974. *Bulletin of the Pan American Health Organization*, *90*(6), 494–505.

Bailey, J., Measham, C. & Umana, M. (1976). Fertility and contraceptive practice: Bogota, 1964-74. *Studies in family planning*, *7*(9), 249–260.

Barber, J. S. & Axinn, W. G. (2004). New ideas and fertility limitation: The role of mass media. *Journal of Marriage and Family*, *66*(5), 1180–1200.

Barone, G. & Mocetti, S. (2021). Intergenerational mobility in the very long run: Florence 1427–2011. *The Review of Economic Studies*, *88*(4), 1863–1891.

Batyra, E. (2016). Fertility and the changing pattern of the timing of childbearing in Colombia. *Demographic Research*, *35*(1), 1343–1372.

Batyra, E. (2020). Increasing educational disparities in the timing of motherhood in the Andean region: A cohort perspective. *Population Research and Policy Review*, *39*(2), 283–309.

Baudin, T., De La Croix, D. & Gobbi, P. (2015). Fertility and childlessness in the

United States. *American Economic Review*, *105*(6), 1852–82.

Baudin, T., De la Croix, D. & Gobbi, P. (2020). Endogenous childlessness and stages of development. *Journal of the European Economic Association*, *18*(1), 83–133.

Beach, B. & Hanlon, W. W. (2019). *Censorship, family planning, and the historical fertility transition* (Working Paper No. 25752). NBER: National Bureau of Economic Research.

Becker, G. (1960). An economic analysis of fertility. In A. J. Coale (Ed.), *Demographic and economic change in developed countries* (Vol. 11, p. 209–231).

Becker, G. (1993). *A treatise on the family* (First Harvard University Press paperback edition ed.). Cambridge, Massachusetts: Harvard University Press.

Becker, G. & Lewis, G. (1973). On the interaction between the quantity and quality of children. *Journal of Political Economy*, *81*(2), S279–S288.

Becker, S., Cinnirella, F. & Woessmann, L. (2010). The trade-off between fertility and education: Evidence from before the demographic transition. *Journal of Economic Growth*, *15*(3), 177–204.

Behrman, J. (2015). Does schooling affect women's desired fertility? Evidence from Malawi, Uganda, and Ethiopia. *Demography*, *52*(3), 787–809.

Behrman, J., Gaviria, A. & Szekely, M. (2001). Intergenerational Mobility in Latin America. *Economia*, *2*(1), 1–31.

Behrman, J., Gaviria, A. & Székely, M. (2003). *Who's in and who's out: social exclusion in Latin America*. Inter-American Development Bank.

Bell, P. L. (1921). *Colombia: A commercial and industrial handbook*. US Government Printing Office.

Beltran Tapia, F. & de Miguel Salanova, S. (2021). Class, literacy and social mobility: Madrid, 1880–1905. *The History of the Family*, *26*(1), 149–172.

Berger, T., Engzell, P., Eriksson, B. & Molinder, J. (2021). *Social mobility in Sweden before the welfare state* (Working Paper No. DP16595). CEPR: Centre for Eco-

nomic Policy Research.

Bértola, L. & Ocampo, J. A. (2012). *The economic development of Latin America since independence*. OUP Oxford.

Bértola, L. & Williamson, J. (2017). *Has Latin American Inequality Changed Direction?: Looking Over the Long Run*. Springer Nature.

Black, S. & Devereux, P. (2011). Recent Developments in Intergenerational Mobility. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (1st ed., Vol. 4, p. 1487-1541). Elsevier.

Blanc, G. (2020). *Modernization before industrialization: cultural roots of the demographic transition in france* (Working Paper No. 3702670). SSRN.

Blanc, G. & Wacziarg, R. (2020). Change and persistence in the Age of Modernization: Saint-Germain-d'Anxure, 1730–1895. *Explorations in Economic History*, *78*, 101352.

Blanco Sánchez, J. C. (2018). Desarrollo histórico de la relación entre la museología y la radio en colombia. *Mestría en Museología y Gestión del Patrimonio*.

Bleakley, H. & Lange, F. (2009). Chronic disease burden and the interaction of education, fertility, and growth. *The Review of Economics and Statistics*, *91*(1), 52–65.

Blossfeld, H.-P. (2009). Educational assortative marriage in comparative perspective. *Annual Review of Sociology*, *35*, 513–530.

Blossfeld, H.-P. & Timm, A. (2003). *Who marries whom?: educational systems as marriage markets in modern societies* (Vol. 12). Springer Science & Business Media.

Bonet, J. & Meisel-Roca, A. (2001). La convergencia regional en Colombia: una visión de largo plazo, 1926-1995. In A. Meisel (Ed.), *Regiones, ciudades y crecimiento económico en Colombia* (pp. 11–56). Bogota: Banco de la República de Colombia.

Bongaarts, J., Mauldin, W. P. & Phillips, J. F. (1990). The demographic impact of family planning programs. *Studies in family planning*, *21*(6), 299–310.

Breen, R. & Karlson, K. B. (2014). Education and social mobility: New analytical approaches. *European Sociological Review*, *30*(1), 107–118.

Bro, N. & Mendoza, M. (2021). Surname affinity in Santiago, Chile: A network-based approach that uncovers urban segregation. *PloS one*, *16*(1), e0244372.

Bulmer-Thomas, V. (2003). *The economic history of Latin America since independence*. Cambridge University Press.

Bushnell, D. (1996). *Colombia: una nación a pesar de sí misma*. Planeta.

Caldwell, J. C. (1980). Mass education as a determinant of the timing of fertility decline. *Population and development review*, *6*(2), 225–255.

Callaway, B., Goodman-Bacon, A. & Sant'Anna, P. H. C. (2021). *Difference-in-differences with a continuous treatment.* (Mimeo)

Camacho, A., Messina, J. & Uribe Barrera, J. (2017). *The expansion of higher education in Colombia: Bad students or bad programs?* (Working Paper No. 13). Documento CEDE: Universidad de Los Andes.

Cardenas, J.-C., Fergusson, L. & Garcia-Villegas, M. (2021). *La quinta puerta: de como la educacion en Colombia agudiza las desigualdades en lugar de remediarlas*. Bogota: Ariel.

Carpio, M. A. & Guerrero, M. E. (2021). Did the colonial mita cause a population collapse? What current surnames reveal in Peru. *The Journal of Economic History*, *81*(4), 1015–1051.

Carrizosa, E. (1990). *Indices de los árboles de las genealogías del Nuevo Reino de Granada de Juan Florez de Ocariz* (Instituto Caro y Cuervo, Ed.). Bogota.

Castro, R. & Tapia, J. (2021). Health, childlessness and poverty in Latin American countries. *Cadernos de Saúde Pública*, *37*.

Castro Martin, T. (1995). Women's education and fertility: results from 26 demo-

graphic and health surveys. *Studies in family planning*, *26*(4), 187–202.

Chetty, R., Friedman, J. N., Saez, E., Turner, N. & Yagan, D. (2017). *Mobility report cards: The role of colleges in intergenerational mobility* (Working Paper No. 23618). NBER: National Bureau of Economic Research.

Chetty, R., Hendren, N., Kline, P. & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, *129*(4), 1553–1623.

Chiappori, P.-A., Dias, M. C. & Meghir, C. (2020). *Changes in assortative matching: Theory and evidence for the US* (Working Paper No. 26932). NBER: National Bureau of Economic Research.

Cho, L.-J., Retherford, R. D. & Choe, M. K. (1986). *The own-children method of fertility estimation*. Population Institute Honolulu, HI.

Cilliers, J. & Fourie, J. (2018). Occupational Mobility during South Africa's Industrial Take-Off. *South African Journal of Economics*, *86*(1), 3–22.

Clark, G. (2007). *A farewell to alms*. Princeton University Press.

Clark, G. (2014). *The son also rises: surnames and the history of social mobility*. Princeton University Press.

Clark, G. & Cummins, N. (2013). *Surnames and social mobility: England 1230-2012* (Working Paper No. 181). Department of Economic History: The London School of Economics.

Clark, G. & Cummins, N. (2014). Intergenerational mobility in England, 1858-2012. Wealth, surnames, and social mobility. *The Economic Journal*, *125*(582), 61–85.

Clark, G., Cummins, N., Hao, Y. & Diaz-Vidal, D. (2015). Surnames: A new source for the history of social mobility. *Explorations in Economic History*, *55*(1), 3–24.

Cleland, J., John; Hobcraft. (1985). *Reproductive change in developing countries:*

Insights from the world fertility survey. Oxford University Press.

Coale, A. J. & Watkins, S. C. (1986). *The Decline of Fertility in Europe*. Princeton University Press.

Coatsworth, J. H. (1993). La independencia Latinoamericana: hipótesis sobre sus costes y beneficios. In L. Prados de la Escosura & S. Amaral (Eds.), *La independencia Americana: consecuencias económicas* (pp. 17–30). Alianza.

Colleran, H., Jasienska, G., Nenko, I., Galbarczyk, A. & Mace, R. (2014). Community-level education accelerates the cultural evolution of fertility decline. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1779), 2–8.

Colmenares, G. (2007). La Formación de la Economía Colonial (1500 – 1740). In *Historia económica de Colombia* (pp. 13–47). Editorial Planeta.

Correa, J. S. (2009). Banca y región en Colombia, 1850-1880. *Revista de Economía Institucional*, *11*(20), 15–44.

Cortés, D., Gallego, J. & Maldonado, D. (2016). On the design of educational conditional cash transfer programs and their impact on non-education outcomes: the case of teenage pregnancy. *The BE Journal of Economic Analysis & Policy*, *16*(1), 219–258.

Dahan, M. & Gaviria, A. (1999). *Sibling Correlations and Social Mobility in Latin America* (Working Paper No. No. 395). IDB: Inter-American Development Bank.

Dahan, M. & Gaviria, A. (2001). Sibling correlations and intergenerational mobility in Latin America. *Economic Development and Cultural Change*, *49*(3), 537–554.

DANE. (1978). *Boletin mensual de estadistica* (No. 322).

Daudin, G., Franck, R. & Rapoport, H. (2019). Can internal migration foster the convergence in regional fertility rates? Evidence from 19th century France. *The Economic Journal*, *129*(620), 1618–1692.

Davila-Ortiz, A. (2010). *El jockey club que yo conoci*. Bogota: Villegas Editores.

De La Croix, D. & Doepke, M. (2003). Inequality and growth: why differential fertility matters. *American Economic Review*, *93*(4), 1091–1113.

De La Croix, D., Schneider, E. B. & Weisdorf, J. (2019). Childlessness, celibacy and net fertility in pre-industrial england: the middle-class evolutionary advantage. *Journal of Economic Growth*, *24*(3), 223–256.

Dell, M. (2010). The persistent effects of Peru's mining mita. *Econometrica*, *78*(6), 1863–1903.

Dellavigna, S. & Gentzkow, M. (2010). Persuasion: Empirical evidence. *Annual Review of Economics*, *2*(1), 643–669.

DellaVigna, S. & La Ferrara, E. (2015). *Economic and social impacts of the media* (Vol. 1). Elsevier.

Dennison, T. & Ogilvie, S. (2014). Does the European marriage pattern explain economic growth? *The Journal of Economic History*, *74*(3), 651–693.

Diaz Vidal, D. (2014). *A surname analysis of social mobility and assortative mating in Chile and Costa Rica.* (Unpublished doctoral dissertation). University of California Davis.

Dobado Gonzalez, R. & Garcia Montero, H. (2010). Colonial origins of inequality in hispanic America? Some reflections based on new empirical evidence. *Revista de Historia Económica - Journal of Iberian and Latin American Economic History*, *28*(2), 253–277.

Doepke, M., Hannusch, A., Kindermann, F. & Tertilt, M. (2022). *The economics of fertility: A new era* (Working Paper No. 29948). NBER: National Bureau of Economic Research.

Dribe, M. & Scalone, F. (2014). Social class and net fertility before, during, and after the demographic transition: A micro-level analysis of Sweden 1880-1970. *Demographic Research*, *30*(1), 429–464.

Duarte, J., Bos, M. S. & Moreno, M. (2012). *Calidad, igualdad y equidad en la*

educación colombiana (análisis de la prueba saber 2009) (Tech. Rep. No. 396). Notas Técnicas: Banco Interamericano de Desarrollo.

Dubuc, S. (2009). Application of the Own-Child method for estimating fertility of women by ethnic groups in the UK. *Journal of Population Research*, *26*, 2007–2225.

Easterlin, R. A. & Crimmins, E. M. (1985). *The fertility revolution: A supply-demand analysis*. University of Chicago Press.

Ermisch, J., Francesconi, M. & Siedler, T. (2006). Intergenerational mobility and marital sorting. *The Economic Journal*, *116*(513), 659–679.

Esteve, A., Saavedra, A. C., López-Colás, J., López-Gay, A. & Lesthaeghe, R. J. (2016). The boom of cohabitation in Colombia and in the Andean region: Social and spatial patterns. In A. Esteve & R. Lesthaeghe (Eds.), *Cohabitation and marriage in the Americas: Geo-historical legacies and new trends* (pp. 187–215). Springer.

Faguet, J.-P., Matajira, C. & Sánchez, F. (2017). *Is extraction bad? encomienda and development in Colombia since 1560* (Working Paper No. 48). Documento CEDE: Universidad de Los Andes.

Feigenbaum, J. J. (2018). Multiple measures of historical intergenerational mobility: Iowa 1915 to 1940. *The Economic Journal*, *128*(612), F446–F481.

Fergusson, L. & Flórez, S. (2021a). Desigualdad Educativa en Colombia. In J. C. Cárdenas, L. Fergusson & M. García-Villegas (Eds.), *La Quinta Puerta: De cómo la educación en Colombia agudiza las desigualdades en lugar de remediarlas* (pp. 81–114). Bogota: Ariel.

Fergusson, L. & Flórez, S. (2021b). Distinción escolar. In J. C. Cárdenas, L. Fergusson & M. García-Villegas (Eds.), *La Quinta Puerta: De cómo la educación en Colombia agudiza las desigualdades en lugar de remediarlas* (pp. 81–114). Bogota: Ariel.

Fergusson, L., Molina, C., Robinson, J. & Vargas, J. F. (2017). *The long shadow of the past: political economy of regional inequality in Colombia* (Working Paper No. 22). Documento CEDE: Universidad de Los Andes.

Fernihough, A. (2017). Human capital and the quantity–quality trade-off during the demographic transition. *Journal of Economic Growth*, *22*(1), 35–65.

Ferreira, L. & Straubhaar, J. (1988). Radio and the new Colombia. *Journal of Popular Culture*, *22*(1), 131.

Flórez, C. E. N. (1992). The fertility transition and family urban income distribution in Colombia. *Sociological Inquiry*, *62*(2), 169–184.

Flórez-Nieto, C. E. (2000). *Las transformaciones sociodemográficas en Colombia durante el siglo XX*. Banco de la Republica de Colombia.

Frye, M. & Lopus, S. (2018). From privilege to prevalence: Contextual effects of women's schooling on African marital timing. *Demography*, *55*(6), 2371–2394.

Fuentes-Vásquez, M. J. (2019). Educational disparities in Colombia 1904-58: New evidence from a regional level approach. *Revista de Historia Económica / Journal of Iberian and Latin American Economic History*, *37*(3), 1–36.

Fuentes-Vásquez, M. J. (2021). *The rise of mass education in Colombia in the first half of the twentieth century* (Unpublished doctoral dissertation). Universitat de Barcelona.

Gabbuti, G. (2022). *When The Leopard Meets Gatsby. Inheritance, Inequality, and Social Mobility in Liberal and Fascist Italy* (Unpublished doctoral dissertation). University of Oxford.

Gallup, J. L., Gaviria, A. & Lora, E. (2003). *Is geography destiny?: lessons from Latin America*. Stanford University Press.

Galor, O. (2005). From Stagnation to Growth: Unified Growth Theory. In P. Aghion & S. Durlauf (Eds.), *Handbook of economic growth* (Vol. 1, pp. 171–293). Elsevier.

Galor, O. & Weil, D. N. (2000). Population, Technology, and Growth: From Maithusian Stagnation to the Demographic Transition and Beyond. *American Economic Review*, *90*(4), 806–828.

Galvis, L. A. & Meisel, A. (2014). *Aspectos regionales de la movilidad social y la igualdad de oportunidades en Colombia* (Working Paper No. 196). Documentos de trabajo sobre economia regional: Banco de la Republica.

García Jaramillo, S., Rodríguez Orgales, C., Sánchez Torres, F. J. & Bedoya Ospina, J. G. (2015). *La lotería de la cuna: la movilidad social a través de la educación en los municipios de Colombia* (Working Paper No. 31). Documento CEDE: Universidad de los Andes.

García-Jimeno, C. (2005). *Colonial institutions and long-run economic performance in Colombia: is there evidence of persistence?* (Working Paper No. 7967). Documento CEDE: Universidad de los Andes.

Garrett, E., Reid, A., Schürer, K. & Szreter, S. (2001). *Changing family size in england and wales: Place, class and demography, 1891–1911*. Cambridge University Press.

González-Bailón, S. & Murphy, T. E. (2013). The effects of social interactions on fertility decline in nineteenth-century France: An agent-based simulation experiment. *Population studies*, *67*(2), 135–155.

Grabill, W. R. & Cho, L. J. (1965). Methodology for the measurement of current fertility from population data on young children. *Demography*, *2*(1), 50–73.

Grafe, R. & Irigoin, M. A. (2006). The Spanish Empire and its legacy: fiscal redistribution and political conflict in colonial and post-colonial Spanish America. *Journal of Global History*, *1*(2), 241–267.

Grant, M. J. & Hallman, K. K. (2008). Pregnancy-related school dropout and prior school performance in KwaZulu-Natal, South Africa. *Studies in family planning*, *39*(4), 369–382.

Guell, M., Mora, J. V. & Telmer, C. I. (2013). The informational content of surnames, the evolution of intergenerational mobility, and assortative mating. *Review of Economic Studies*, *82*(2), 693–735.

Guell, M., Rodríguez Mora, J. V. & Telmer, C. I. (2007). *Intergenerational Mobility and the Informative Content of Surnames* (Working Paper No. DP6316). CEPR: Centre for Economic Policy Research.

Guillén de Iriarte, M. C. (2006). *Los estudiantes del Colegio Mayor de Nuestra Señora del Rosario, 1773-1826*. Bogota: Universidad del Rosario.

Guillén de Iriarte, M. C. (2008). *Los estudiantes del Colegio Mayor de Nuestra Señora del Rosario, 1826-1842*. Bogota: Universidad del Rosario.

Guinnane, T. W. (2011). The historical fertility transition: A guide for economists. *Journal of Economic Literature*, *49*(3), 589–614.

Guinnane, T. W., Okun, B. S. & Trussell, J. (1994). What do we know about the timing of fertility transitions in Europe? *Demography*, *31*(1), 1–20.

Gutiérrez de Pineda, V. (1968). *Familia y cultura en Colombia. Tipologías, funciones y dinámica de la familia. Manifestaciones múltiples a través del mosaico cultural y sus estructuras sociales*. Universidad Nacional de Colombia, Tercer Mundo Editores, Bogotá.

Guzman, J. M., Singh, S., Rodriguez, G. & Pantelides, E. (1996). *The fertility transition in Latin America*. Oxford University Press.

Heaton, T. B. & Forste, R. (1998). Education as policy: The impact of education on marriage, contraception, and fertility in Colombia, Peru, and Bolivia. *Social biology*, *45*(3-4), 194–213.

Helg, A. (2001). *La educación en Colombia, 1918-1957: una historia social, económica y política*. Universidad Pedagógica Nacional.

Helg, A. (2014). La limpieza de sangre bajo las reformas borbónicas y su impacto en el Caribe neogranadino. *Boletín de historia y antigüedades*, *101*(858), 143–180.

Heligman, L. & Pollard, J. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, *107*(1), 49–80.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N. & Verashchagina, A. (2007). The inheritance of educational inequality: international comparisons and fifty-Year trends. *The B.E. Journal of Economic Analysis & Policy*, *7*(2), 48.

Hufford, G. A. (2002). The ITS irregular terrain model, version 1.2. 2 the algorithm. *Institute for Telecommunication Sciences, National Telecommunications and Information Administration, US Department of Commerce.*.

Instituto Colombiano de Cultura Hispanica. (1992). *Geografia humana de Colombia. Region del Pacifico - Tomo IX*. Bogota: Instituto Colombiano de Antropología e Historia.

Iregui-Bohórquez, A. M., Melo-Becerra, L. A., Ramírez-Giraldo, M. T. & Tribín-Uribe, A. M. (2020). *The path to gender equality in Colombia: Are we there yet?* (Working Paper No. 1131). Borradores de Economía: Banco de la República de Colombia.

Jaeger, D. A., Joyce, T. J. & Kaestner, R. (2020). A cautionary tale of evaluating identifying assumptions: did reality tv really cause a decline in teenage childbearing? *Journal of Business & Economic Statistics*, *38*(2), 317–326.

Jaramillo-Echeverri, J., Álvarez, A. & Bro, N. (2021). *Surnames and social rank: Long-term traits of social mobility in colombia and chile* (Working Paper No. 01). CAF: Development Bank of Latin America.

Jaramillo-Echeverri, J., Meisel-Roca, A. & Ramírez-Giraldo, M. T. (2019). More than 100 years of improvements in living standards: the case of Colombia. *Cliometrica*, *13*(3), 323–366.

Jaramillo Uribe, J. (1963). Esclavos y señores de la sociedad colombiana del siglo XVIII. *Anuario Colombiano de Historia Social y de la Cultura*, *1*, 3–62.

Jaramillo Uribe, J. (1964). La población indígena de Colombia en el momento de la conquista y sus transformaciones posteriores: primera parte. *Anuario colombiano de historia social y de la cultura*(2), 239–293.

Jayne, S. H. & Guilkey, D. K. (1998). Contraceptive determinants in three leading countries. *Population Research and Policy Review*, *17*(4), 329–350.

Jejeebhoy, S. J. (1995). Women's education, autonomy, and reproductive behaviour: Experience from developing countries. *OUP Catalogue*, *22*(4).

Jensen, R. (2012). Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India. *The Quarterly Journal of Economics*, *127*(2), 753-792.

Kalmanovitz, S. (1983). Los orígenes de la industrialización en colombia: 1890-1929. *Cuadernos de economía*, *5*(5), 79–126.

Kalmanovitz, S. (2001). *Las instituciones y el desarrollo económico en Colombia*. Bogota: Norma.

Kalmanovitz, S. (2008). *La economía de la Nueva Granada*. Universidad de Bogotá Jorge Tadeo Lozano, Ciencias Económico-Administrativas.

Kalmanovitz, S. (2015). *Breve historia económica de Colombia*. Universidad de Bogotá Jorge Tadeo Lozano.

Kearney, M. S. & Levine, P. B. (2015). Media influences on social outcomes: The impact of MTV's 16 and pregnant on teen childbearing. *American Economic Review*, *105*(12), 3597–3632.

King, T. E. & Jobling, M. A. (2009). What's in a name? y chromosomes, surnames and the genetic genealogy revolution. *Trends in genetics*, *25*(8), 351–360.

Knodel, A. J. (2015). *The decline of fertility in Germany, 1871-1939*. Princeton University Press.

Komlos, J. (1998). Shrinking in a growing economy? the mystery of physical stature during the industrial revolution. *The Journal of Economic History*, *58*(3), 779–

802.

Kostaki, A. & Panousis, V. (2001). Expanding an abridged life table. *Demographic Research*, *5*, 1–22.

Kravdal, Ø. (2002). Education and fertility in sub-Saharan Africa: Individual and community effects. *Demography*, *39*(2), 233–250.

La Ferrara, E., Chong, A. & Duryea, S. (2012). Soap operas and fertility: Evidence from Brazil. *American Economic Journal: Applied Economics*, *4*(4), 1–31.

Lindo, J. M., Myers, C. K., Schlosser, A. & Cunningham, S. (2020). How far is too far? new evidence on abortion clinic closures, access, and abortions. *Journal of Human Resources*, *55*(4), 1137–1160.

Liu, D. H. & Raftery, A. E. (2020). How do education and family planning accelerate fertility decline? *Population and Development Review*, *46*(3), 409–441.

Livi-Bacci, M. (2017). Social-group forerunners of fertility control in Europe. In A. Coale & S. Watkins (Eds.), *The decline of fertility in Europe* (pp. 182–200). Princeton University Press.

Londoño, J. L. (1995). *Distribucion del ingreso y desarrollo economico: Colombia en el siglo XX* (T. M. Editores, Ed.). Bogota.

Long, J. & Ferrie, J. (2013). Intergenerational occupational mobility in Great Britain and the United States since 1850. *American Economic Review*, *103*(4), 1109–37.

Long, J. & Ferrie, J. (2018). Grandfathers matter(ed): Occupational mobility across three generations in the US and Britain, 1850–1911. *Economic Journal*, *128*(612), F422–F445.

López Toro, A. (1968). *Análisis demográfico de los censos colombianos, 1951 y 1964*. Ediciones Universidad de los Andes.

Maloney, W. F. & Valencia Caicedo, F. (2016). The persistence of (subnational) fortune. *The Economic Journal*, *126*(598), 2363–2401.

Malthus, T. (1803). *An essay on the principle of population.* Cambridge University Press.

Marshall, J. & Calderon, V. (2006). *Social exclusion in education in Latin America and the Caribbean* (Working Paper). Sustainable Development Department: Inter-American Development Bank.

Marteleto, L., Lam, D. & Ranchhod, V. (2008). Sexual behavior, pregnancy, and schooling among young people in urban South Africa. *Studies in family planning*, *39*(4), 351–368.

Mateos, P., Longley, P. A. & O'Sullivan, D. (2011). Ethnicity and population structure in personal naming networks. *PloS one*, *6*(9), e22943.

McSweeney, K. & Arps, S. (2005). A "Demographic Turnaround": The rapid growth of indigenous populations in lowland Latin America. *Latin American Research Review*, *40*(1), 3–29.

Meisel, A. & Ramírez, M. T. (2010). *Economía colombiana del siglo XIX.* Fondo de Cultura Económica - Banco de la República.

Meisel-Roca, A., Ramírez-Giraldo, M. T. & Jaramillo-Echeverri, J. (2016). Too late but profitable: Railroads in Colombia during 1920–1950. *Investigaciones de Historia Económica-Economic History Research*, *12*(3), 165–180.

Mejía, D., Ramírez-Giraldo, M. T. & Tamayo-Castaño, J. A. (2008). *The demographic transition in Colombia: theory and evidence* (Working Paper No. 538). Borradores de Economía: Banco de la República.

Mejia, J. (2012). *Diccionario biográfico y genealógico de la élite antioqueña y viejocaldense. Segunda mitad del siglo XIX y primera del XX.* Pereira: Editorial Red Alma Mater.

Mejía, W. J. (1996). *Real Colegio Mayor y Seminario de San Bartolomé: nobleza e hidalguía; colegiales de 1605 a 1820.* Instituto Colombiano de Cultura Hispánica.

Melo, J. O. (1996). *Historiografía colombiana: realidades y perspectivas.* Seduca.

Melo, J. O. (2017). *Historia mínima de Colombia.* El Colegio de Mexico AC.

Mendoza-Hoyos, H. (1968). Research studies on abortion and family planning in Colombia. *The Milbank Memorial Fund Quarterly*, *46*(3), 223–236.

Miller, G. (2010). Contraception as development? New evidence from family planning in Colombia. *The Economic Journal*, *120*(545), 709–736.

Miller, G. & Babiarz, K. S. (2016). Family planning program effects: Evidence from microdata. *Population and Development Review*, *42*(1), 7–26.

Miller, G. & Urdinola, B. P. (2010). Cyclicality, mortality, and the value of time: The case of coffee price fluctuations and child survival in Colombia. *Journal of Political Economy*, *118*(1), 113–155.

Minnesota Population Center. (2019). Integrated public use microdata series, international: Version 7.2 [dataset]. *Minneapolis: University of Minnesota.*

Modalsli, J. (2015). Estimating occupational mobility with covariates. *Economics Letters*, *133*(804), 77–80.

Mora Bustamante, A. (2016). *Consecuencias a largo plazo de la encomienda en Colombia: persistencia institucional.* (Mimeo)

Mosquera, S. A. (2014). *Afrochochoanos: orígenes y troncos familiares.* Quibdó: Muntú - Bantú: Fundación Social Afrocolombiana.

Moyano, L. M. & Galvis, L. A. (2014). *¿Oportunidades para el futuro?: la movilidad social de los adolescentes en Colombia* (Working Paper No. 211). Documentos de trabajo sobre economia regional: Banco de la Republica.

Narayan, A., Van der Weide, R., Cojocaru, A., Lakner, C., Redaelli, S., Gerszon Mahler, D., ... Thewissen, S. (2018). *Fair Progress?: Economic Mobility across Generations around the World.* The World Bank.

Nieto Arteta, L. E. (1999). *El café en la sociedad colombiana.* El Áncora.

Novotnỳ, J. & Cheshire, J. A. (2012). The surname space of the Czech Republic:

examining population structure by network analysis of spatial co-occurrence of surnames. *PloS one*, *7*(10), e48568.

Ocampo, J. A. (2004). Latin america's growth and equity frustrations during structural reforms. *Journal of Economic Perspectives*, *18*(2), 67–88.

O'Connell, M. (1981). Regional fertility patterns in the United States: convergence or divergence? *International Regional Science Review*, *6*(1), 1–14.

OECD. (2018). *A Broken Social Elevator? How to Promote Social Mobility*. Paris: OECD Publishing.

Olarte, R. S. (1993). Estructuración y disolución de un sistema de reclutamiento de elites: El caso de Santafé de Bogotá, en Nueva Granada. *Boletín Cultural y Bibliográfico*, *30*(32), 3-23.

Olivetti, C. & Paserman, M. D. (2015). In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850-1940. *American Economic Review*, *105*(8), 2695–2724.

Olken, B. A. (2009). Do television and radio destroy social capital? Evidence from Indonesian villages. *American Economic Journal: Applied Economics*, *1*(4), 1–33.

Osili, U. O. & Long, B. T. (2008). Does female schooling reduce fertility? Evidence from Nigeria. *Journal of Development Economics*, *87*(1), 57–75.

Ott, E. R. (1977). Population policy formation in Colombia: The role of ASCOFAME. *Studies in family planning*, *8*(1), 2–10.

Palacios, M. A. & Safford, F. (2012). *Historia de Colombia. País fragmentado, sociedad dividida*. Universidad de los Andes, Facultad de Administración.

Pandit, K. & Bagchi-Sen, S. (1993). The spatial dynamics of us fertility, 1970-1990. *Growth and change*, *24*(2), 229–246.

Parr, N. (2002). Family planning promotion, contraceptive use and fertility decline in Ghana. *African Population Studies*, *17*(1), 83–101.

Parrado, E. (2000). Social change, population policies, and fertility decline in Colombia and Venezuela. *Population Research and Policy Review*, *19*(5), 421–457.

Parry, M. (2013). *Broadcasting birth control*. Rutgers University Press.

Parson, J. (1950). *La colonización antioqueña en el occidente de Colombia*. Imprenta Departamental de Antioquia, Medellín.

Pearce, J. & Montoya, J. D. V. (2022). *Elites, poder y principios de dominacion en Colombia (1991-2022)* (Working Paper). LACC: London School of Economics.

Pérez, S. (2017). The (South) American dream: mobility and economic outcomes of first-and second-generation immigrants in nineteenth-century Argentina. *The Journal of Economic History*, *77*(4), 971–1006.

Pérez, S. (2019). Intergenerational occupational mobility across three continents. *The Journal of Economic History*, *79*(2), 383–416.

Peyser, A. & Chackiel, J. (1993). *La población indigena en los censos de América Latina.*

Piñeros-Petersen, M. & Ruiz-Salguero, M. (1998). Aspectos demográficos en comunidades indígenas de tres regiones de Colombia. *Salud Pública de México*, *40*(4), 324–329.

Pitt, M. M., Rosenzweig, M. R. & Gibbons, D. M. (1993). The determinants and consequences of the placement of government programs in Indonesia. *The World Bank Economic Review*, *7*(3), 319–348.

Poston, D. L., Kramer, K. B., Trent, K. & Yu, M.-Y. (1983). Estimating voluntary and involuntary childlessness in the developing countries. *Journal of Biosocial Science*, *15*(4), 441–452.

Potter, J. & Ordoñez, M. (1976). The completeness of enumeration in the 1973 census of the population of Colombia. *Population Index*, *42*(3), 377–403.

Potter, J., Ordoñez, M. & Measham, A. (1976). The rapid decline in Colombian fertility. *Population and Development Review*, *2*(3), 509–528.

Prada, E. & Ojeda, G. (1987). Selected findings from the demographic and health survey in Colombia, 1986. *International Family Planning Perspectives*, *13*(4), 116–120.

Prada-Salas, E. (1996). The fertility transition and adolescent childbearing: the case of Colombia. In J. M. Guzman, S. Singh, G. Rodriguez & E. Pantelides (Eds.), *The fertility transition in Latin America* (pp. 310–322). Oxford University Press.

Prados De La Escosura, L. (2005). *Growth, inequality, and poverty in Latin America: historical evidence, controlled conjectures* (Working Paper No. 05-41(04)). Economic History Department: Universidad Carlos III.

Prados de la Escosura, L. (2009). Lost decades? Economic performance in post-independence Latin America. *Journal of Latin American Studies*, *41*(2), 279–307.

Ramírez-Giraldo, M. T. & Téllez-Corredor, J. P. (2006). *La educación primaria y secundaria en Colombia en el siglo XX* (Working Paper No. 379). Borradores de Economía: Banco de la Republica.

Ramirez-Zuluaga, J. S. (2016). *Movilidad social intergeneracional por ingresos en Colombia* (Unpublished doctoral dissertation). Universidad Nacional de Colombia.

Reid, A., Jaadla, H., Garrett, E. & Schürer, K. (2020). Adapting the own children method to allow comparison of fertility between populations with different marriage regimes. *Population Studies*, *74*(2), 197–218.

Requena, M. (1968). The problem of induced abortion in Latin America. *Demography*, *5*(2), 785–799.

Ripoll Echeverría, M. T. (2006). *La élite en Cartagena y su tránsito a la República-revolución política sin renovación social*. Universidad de los Andes, Facultad de Ciencias Sociales.

Robinson, J. & Urrutia, M. (2007). *Economía colombiana del siglo XX: Un análisis cuantitativo*. Fondo de cultura Económica - Banco de la República.

Roldán, M. (2014). Acción Cultural Popular, responsible procreation, and the roots of social activism in rural Colombia. *Latin American Research Review*, 27–44.

Rosero-Bixby, L., Castro-Martín, T. & Martín-García, T. (2009). Is Latin America starting to retreat from early and universal childbearing? *Demographic Research*, *20*, 169–194.

Santavirta, T. & Stuhler, J. (2020). *Name-Based estimators of intergenerational mobility: Evidence from Finnish veterans.* (Mimeo)

Schwartz, C. R. (2013). Trends and variation in assortative mating: Causes and consequences. *Annual Review of Sociology*, *39*, 451–470.

Shea, G. A., Burch, T. K., Pérez, G., Ordoñez, M., Van Kemanade, J., Hutjes, J. & Hellegers, A. E. (1971). Catholic parish priests and birth control: A comparative study of opinion in Colombia, the United States, and the Netherlands. *Studies in Family Planning*, *2*(6), 121–136.

Simmons, A. & Cardona, R. (1974). Colombia: Stages of family planning adoption. *Studies in Family Planning*, *5*(2), 42–49.

Sobek, M. & Kennedy, S. (2009). *The development of family interrelationship variables for international census data* (Working Paper No. 02). Minnesota Population Center: Minneapolis: University of Minnesota.

Sokoloff, K. L. & Engerman, S. L. (2000). Institutions, factor endowments, and paths of development in the New World. *Journal of Economic perspectives*, *14*(3), 217–232.

Solon, G. (2018). What do we know so far about multigenerational mobility. *The Economic Journal*, *128*(612), 1–13.

Spolaore, E. & Wacziarg, R. (2022). Fertility and modernity. *The Economic Journal*, *132*(642), 796–833.

Stycos, M. & Avery, R. (1975). Family Planning via the airwaves: radio campaign in Colombia. In *The clinic and information flow* (pp. 151–164). London: Lexington Books.

Tan, H. R. (2019). More is less? The impact of family size on education outcomes in the United States, 1850–1940. *Journal of Human Resources*, *54*(4), 1154–1181.

Thiede, B. C. & Gray, C. (2020). Characterizing the indigenous forest peoples of Latin America: Results from census data. *World Development*, *125*, 104685.

Tienda, M., Diaz, V. G. & Smith, S. A. (1985). Community education and differential fertility in Peru. *Canadian Studies in Population*, *12*(2), 137–158.

Timæus, I. M. (2021). The Own-Children Method of fertility estimation: The devil is in the detail. *Demographic Research*, *45*(May 2022), 825–840.

Tirado Mejía, A. (1998). *Introducción a la historia económica de Colombia.* Universidad de Antioquia.

Torche, F. (2014). Intergenerational mobility and inequality: The Latin American case. *Annual Review of Sociology*, *40*, 619–642.

Torche, F. & Corvalan, A. (2018). Estimating intergenerational mobility with grouped data: A critique of Clark's the Son Also Rises. *Sociological Methods and Research*, *47*(4), 787–811.

Tovar, J. (2007). *La manumisión en Colombia: 1821-1851. Un análisis cuantitativo* (Working Paper No. 28). Documento CEDE: Universidad de Los Andes.

Tovar, J. A. & Tovar, H. (2009). *El oscuro camino de la libertad. Los esclavos en Colombia, (1821-1851).* Bogota: Ediciones Uniandes.

Udry, J. R., Clark, L. T., Chase, C. L. & Levy, M. (1972). Can mass media advertising increase contraceptive use? *Family Planning Perspectives*, *4*(3), 37–44.

Universidad de los Andes. (2022). *Resultados Generales de Encuesta Laboral.* (Unpublished)

Urdinola, B. P. & Ospino, C. (2015). Long-term consequences of adolescent fertility:

The Colombian case. *Demographic Research*, *32*, 1487–1518.

Urrutia, M. & Berry, A. (1975). *La distribución del ingreso en Colombia.* La Carreta, Medellín.

Valencia Caicedo, F. (2019). The mission: Human capital transmission, economic persistence, and culture in South America. *The Quarterly Journal of Economics*, *134*(1), 507–556.

Vejarano, F. & McCaa, R. (2002). *Homologacion de los microdatos censales colombianos, 1964-1993* (DANE, P. C. U. of Minnesota & CIDS, Eds.).

Velasquez Murillo, R. (1962). Gentilicios africanos del occidente de Colombia. *Revista colombiana de folclor*, *II*(7), 107–149.

Villalobos, Á. R. A. (2005). La limpieza de sangre en el Colegio de Abogados de Caracas a finales del siglo XVIII. *Fronteras de la Historia*(10), 305–328.

Voigtländer, N. & Voth, H.-J. (2006). Why England? Demographic factors, structural change and physical capital accumulation during the Industrial Revolution. *Journal of economic growth*, *11*(4), 319–361.

Vosters, K. (2018). Is the simple law of mobility really a law? Testing Clark's hypothesis. *The Economic Journal*, *128*(612), F404–F421.

Wang, T. (2021). Media, pulpit, and populist persuasion: evidence from father coughlin. *American Economic Review*, *111*(9), 3064–92.

Ward, Z. (2021). *Intergenerational mobility in American history: Accounting for race and measurement error* (Working Paper No. 29256). NBER: National Bureau of Economic Research.

Weinberger, M. B. (1987). The relationship between women's education and fertility: selected findings from the world fertility surveys. *International Family Planning Perspectives*, *13*(2), 35–46.

Westoff, C. F. & Rodriguez, G. (1995). The mass media and family planning in Kenya. *International Family Planning Perspectives*, *21*(1), 26.

Williams, J. S. & Singh, B. K. (1976). The impact of mass media on fertility rates: An exploration. *Journal of Applied Communication Research*, *4*(1), 19–24.

Williamson, J. G. (2010). Five centuries of Latin American income inequality. *Journal of Iberian and Latin American Economic History*, *28*(2), 227–252.

Williamson, J. G. (2015). Latin american inequality: Colonial origins, commodity booms or a missed twentieth-century leveling? *Journal of Human Development and Capabilities*, *16*(3), 324-341.

Wills Franco, M. (1976). *Diferencias regionales de la fecundidad en colombia.* Asociación Colombiana de Facultades de Medicina, División de Medicina Social.

Wilson, C. & Woods, R. (1991). Fertility in england: a long-term perspective. *Population Studies*, *45*(3), 399–415.

Woessmann, L., Ursprung, H. W. & Schuetz, G. (2005). *Education policy and equality of opportunity* (Working Paper No. 5694). Munich: Center for Economic Studies and Ifo Institute (CESifo).

Zapata Olivella, M. & Mina Aragon, W. (2014). *El árbol brujo de la libertad: África en Colombia, orígenes-transculturación-persistencia, ensayo histórico mítico.* Bogota: Ediciones Desde abajo.