PHD-FDEF-2021-019
The Faculty of Law, Economics and Finance

DISSERTATION

Defence held on 14/10/2021 in Luxembourg
to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG
EN SCIENCES ÉCONOMIQUES

By

ANA CECILIA MONTES-VIÑAS
Born on 26 April 1989 in Sincelejo (Colombia)

ESSAYS ON THE ECONOMICS OF INTERNATIONAL MIGRATION

Dissertation defence committee
Dr Michel Beine, dissertation supervisor
Professor, Université du Luxembourg

Dr Arnaud Dupuy, Chairman
Professor, Université du Luxembourg

Dr Skerdiladja Zanaj, Vice Chairman
Professor, Université du Luxembourg

Dr Simone Moriconi, member
Professor, IÉSEG School of Management

Dr Lionel Ragot, member
Professor, Université Paris X Nanterre

ABSTRACT OF THE DISSERTATION

The present dissertation consists of three main chapters of self-contained works about international human migration and migrant's integration in the host society. The first chapter introduces the general outline of the dissertation and briefly explains the research questions explored in each chapter.

The second chapter studies the effect of migration networks and long-term cultural distance on migration flows. The main research question is whether the diaspora effect on migration flows is larger when the cultural distance between the country of origin and destination is large. We use an unbalanced panel database of bilateral migration flows data from 1960 to 2009 for about 175 sending countries to 32 destination countries. We proxy long-term cultural distance using ancestral distance, also called genetic distance. Based on the micro-founded gravity model for migration, we estimate the interaction between the diaspora size and ancestral distance using the Poisson Pseudo Maximum Likelihood (Santos Silva & Tenreyro, 2006). We find evidence of a positive and significant interaction effect between the network effect and ancestral distance on international migration flows, however, this effect is small once we control for omitted unobserved determinants of migration flows.

The third chapter studies the educational performance of the children of migrants in the United States of America. It pushes forward the hypothesis that misalignment between expectations and aspirations crucially affects the educational outcomes of immigrant young adults. It shows that the difference in school performance between migrant children and natives lies within the aspirations and expectations that migrant children form. This chapter uses the National Longitudinal Study of Adolescent to Adult Health (Add Health), a longitudinal database representative of American high schools that surveys adolescents between the

grades 7-12 collected by the Carolina Population Center. The chapter shows that a positive difference between aspirations and expectations is a driving force for higher effort and better education outcomes of immigrant teenagers. This force resolves the well-known immigrant paradox. This result is specific to migrant children and does not hold for second-generation migrant pupils who appear quite acculturated to the USA context.

In the fourth chapter, the relationship between financial aid and foreign education at the postgraduate level is evaluated. I use a sample of Colombian graduates who applied to a financial aid program that sponsors the completion of master's degrees abroad. It studies the case of a Scholarship-Loan program provided by the Colfuturo Foundation and the Administrative Department of Science, Technology, and Innovation (Colciencias). Students with an undergraduate degree can apply to receipt a fund of 50000 dollars to finance their postgraduate studies in any country in the world, with the condition to come back to Columbia. The characteristics of the selection process of the program allow the implementation of a Regression Discontinuity Design (RDD) to estimate the causal effect of the scholarship-loan program on the probability of completing studies abroad. Given Colfuturo's selection process, the assignment of the fund is expected to be distributed quasi-randomly for the applicants around the vicinity of a cut-off point. This methodology allows for the estimation of the local average treatment effect (LATE) of the program. I employ the administrative records from the scholarship provider combined with internet sources. The results show that the scholarship-loan program is an effective tool to increase the probability of completing studies abroad by approximately 30 percentage points. The results are extremely robust across estimation methods.

ACKNOWLEDGEMENTS

This thesis was completed with assistance from a lot of people whom I would like to thank. I want to express my sincere gratitude to my supervisor Prof. Dr. Michel Beine, for his patience, invaluable support, scholarly guidance through all these years, and the trust placed in me to undertake this research. I am incredibly grateful to Prof. Dr. Skerdiladja Zanaj. Thank you for the wisdom, consistent guidance, and tremendous support. I want to use this space to say that you are an excellent role model and mentor. I am deeply grateful to Prof. Dr. Frédéric Docquier for the constructive comments. Moreover, I also thank Prof. Dr. Arnaud Dupuy for the advice provided as a member of my CET and the external members of the board Prof. Dr. Simone Moriconi and Prof. Dr. Lionel Ragoot for their valuable comments.A special thanks go to Prof. Dr. Clemente Forero and Prof. Dr. Michaela Kreyenfeld, two cornerstones in the development of my interest in research.

Moreover, I am indebted to the professors and staff in the Department of Economics and Management of the University of Luxembourg. A big thank you to Roswitha, Marina, Elisa, Anne-Sophie, and Celine for their assistance. Also, I would like to thank Dr. Majlinda Joxhe and Dr. Joel Machado for the warm welcome to the team. To Dr. Clotilde Mahé for her comments and re-views. I like to thank my friends Nijat, Veronique, Nora, Basheer, Huyen, Kakit, Nicole, Andrei, Ruth, Basit, Rosa Lisa, Basit, Maxime, Ksenia, Ni, Silvia, Rana, Gunn Elin, Rebecca, and Sebastian for your great friendship and support. Special thanks go to my friends Catherine and Monica for their guidance, wisdom, and care.

Finally and more importantly, I thank my family for their unfailingly support and unconditional love. My parents Pablo and Cecilia who have taught me the value of hard work and to aim high. To my brother Pablo who has always been by my side.

Dedico este trabajo a mis padres y a mi hermano.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Human migration is not a new phenomenon. The movement from one place to another different from the habitual place of residence is inherent to human beings. Despite being such a natural process, it still ignites political debates and concerns among voters in every country. Many questions arise when we think about migrants from another country, region, or city. Will migrants occupy all the jobs? Will migrants deplete welfare budgets? Are migrants criminals? Will migrants and their children integrate into their host societies? Will migrants' values dilute "our" values? The literature on the economics of international human migration has dedicated many pages to solving these questions using the scientific method. This thesis will not cover all these questions, however, it does center on the analysis of specific myths behind the determinants of the international movement of people and the integration of migrants.

Will more culturally distant migrants arrive at higher rates than less culturally distant ones? are the pre-existing stock of migrants contributing to it? These questions are addressed in chapter 2 of this thesis. Following Paul Collier, who raised this concern in his book on international migration (Collier, 2013), chapter 2 submits these questions to a test using quantitative methods. Cultural differences between destination and origin countries deter people from moving abroad (Adserà and Pytliková, 2015; Belot and Ederveen, 2012). The results presented in chapter 2 confirm that long-term cultural distance, measured using ancestral distance, is associated with lower bilateral migration flows. These cultural differences could affect greatly the integration process of migrants in the host countries. However, pre-existing migrants from the same origin country could make this process easier. In fact, chapter 2 affirms that the pre-existing stock of migrants from further cultural distant countries attract future inflows to a greater extend when compared to the effects of pre-existing stock of migrants from less cultural distant countries. While this effect is statistically significant, its magnitude is rather small after taking into consideration unobserved determinants of international migration flows.

Chapter 3 focuses on the integration of the children of immigrants in the school. Will migrant children fall behind in school? Literature in this subject is vast and it suggests that on average, migrant children perform worse at school when compared to natives (Sweetman and van Ours, 2014; Dustmann et al., 2012). However, in the US and other English-speaking countries, it is the opposite. Migrant children seem to be over-performing when compared to their peers (Palacios et al., 2008; Feliciano and Lanuza, 2017). This paradox is addressed in the third chapter of this thesis. This chapter raises the idea that aspirations and expectations formed by the children of migrants during their early teenage years determine their school outcomes and the discrepancies in performance when compared to their native peers. Moreover, it shows a positive association between the misalignment in aspirations and expectations and the difference in school GPA between migrant teens and native peers. This difference in the high school GPA between migrant and native teenagers is negligible after controlling for an extensive set of variables that relates to the student, the family, and the school. For teens who do not display such misalignment, there is no statistically significant difference in their GPA compared to their peers.

Chapter 4 concentrates on a particular group of migrants: international students. It studies how financial aspects influence the probability of moving abroad to complete higher education at the master's level. This chapter implements a Regression Discontinuity Design

(RDD) to casually estimate the impact of a financial aid program on studying abroad. The present paper contributes to the literature related to the determinants of studying abroad by accounting for selection bias as a result of self-selection into financial aid. The results confirm that being a recipient of Colfuturo's financial help increased the probability of studying abroad at the postgraduate level by 30 percentage points approximately. Nearly 98% of the students in the sample who are eligible for the financial aid move abroad to study compared to 63.4 % who were not eligible to the program. This chapter discusses how easing cost barriers can encourage the movement of young professionals to study in foreign countries.

One of the major challenges in migration research is the availability of high-frequency data that tracks people's mobility. To estimate the inflow of international migrants, the literature traditionally employs the difference in the stock of immigrants from one census period to the other. However, this method measures immigration inflows with some error (Beine et al., 2010a). In chapters 2 and 4 of this thesis, alternative data sources are used. The second chapter makes use of migrations inflows obtained from population registries, border statistics, and/or permits (residence/work) instead of changes in the stock of migrants. Results in chapter 2 show that using these alternative information sources leads to similar estimates of the network effect on international migration flows than when using the differences in migration stocks.

In chapter 4, internet sources and administrative records are used to track international students. While internet resources have not been widely employed to investigate international students, the use of these digital sources of information to track migrants started to emerge. For example, recent literature has used phone call detail records (CDR) to estimate internal migration trends and forced displacement (Blumenstock, 2012; Zealand, 2012; Wilson et al., 2016; Lu et al., 2016; Fiorio et al., 2017; Beine et al., 2021). For example, 1.9 million SIM cards were used to track trends in population movements in Haiti during the 2010 earthquake (Bengtsson et al., 2011). Another strand of literature exploits digital media to understand patterns of migration. For example, using a large extract of Yahoo! e-mails and user's IP addresses, Zagheni and Weber (2012) estimate international migration rates. In a similar fashion, State et al. (2014) estimate trends in international migration of high skilled workers using data retrieved from LinkedIn. Zagheni et al. (2014) use Twitter user's location to infer international and internal migration patterns. While these alternative sources are also subject to measurement error, they complement traditional estimates and can potentially provide interesting insights about the direction of flows and new patterns of migration.[1]

To resume, this dissertation consists of three empirical studies that can be read independently. Each chapter addresses a research question, contains a specific literature review, methodology, empirical results, conclusions, and discussion of the results. As a summary: first, this thesis estimates the joint effects of the pre-existing stocks of migrant and ancestral distance to predict migration inflows. Second, it studies the children of migrants in the United States of America, their performance in high school when compared to the children of natives, and the role of educational aspirations and expectations as determinant factors behind their educational performance. Third, it examines international students and the role of financial aid on the successful completion of higher education abroad.

[1]Data coming from social media could suffer from selection bias problems since it might under-represent certain age groups or populations that do not use social media. Unlike CDR, people can be traced if they move across international borders. CD data might not be suitable to estimate international movement patterns since roaming could be a very expensive service for the average migrant. Another caveat of using mobile phone records is that it also under-represents the total movements since some population groups like the very poor, children or elderly adults who might not own a phone, or provinces that have low mobile phone use or low coverage network. Moreover, the records contain deficient information about the socio-demographic characteristics of the user, making it difficult to understand the potential selection of the bias.

CHAPTER 2
MIGRATION DIASPORAS AND LONG-TERM CULTURAL DISTANCE: TESTING THE COLLIER'S HYPOTHESIS

This paper is a joint work with Prof. Dr. Michel Beine and Prof. Dr. Frédéric Docquier

2.1 Introduction

While the movement of people has been the missing ingredient of globalization in the Twentieth-Century, the increase in the magnitude of migration flows across countries has expanded the interest from policymakers regarding the composition and characteristics of migrants. With more people wishing to cross borders, there has been rising pressure from voters in many countries to introduce more restrictive or selective immigration policies. Understanding international migration patterns has become a crucial component for designing consistent and sound management of foreign immigrants. The identification of key factors influencing the mobility of people across countries and quantifying the most important effects of migration flows are necessary steps to take before designing any policy of that kind.

Many factors have been found to favor the attractiveness of specific destinations for potential migrants, and therefore to raise migration between countries. Among these factors, the existence of migrants communities (migrant networks) at the destination country is among the most important ones. The presence of people originating from one's country or region increases, in an unambiguous way, the probability that a potential migrant will choose that destination. It also raises the likelihood that this migrant will be in that location, conditional on this choice. The so-called network effect generates important dynamic patterns in the mobility process of people. While no one questions its existence nowadays, there is still a lot of debate about its quantification and the factors that influence its magnitude.

This paper addresses this question in a specific way. We look at whether the network effect varies with the cultural distance between the origin country of the migrant and the destination. This question was first raised by Paul Collier in his book on international migration (Collier, 2013). One central thesis described in the book argues that the network effect is higher for migrants coming from more culturally distant countries. This question goes beyond intellectual or academic curiosity. If this idea is confirmed, it implies that for a given size of the diaspora, one particular country will attract more migrants from countries culturally different with respect to the native population. In turn, following Collier (2013) at least, this cultural diversity could have long-run adverse consequences in the destinations countries.[1] With the exception of Collier and Hoeffler (2018), this conjecture has not been evaluated in empirical work. We test it using new data on gross migration flows and long-term cultural distance and some state-of-the-art econometric techniques suited to deal with this type of data.

An essential issue in this empirical task is the choice of particular measures of the cultural distance across countries. While some attempts have been made to capture cultural divergence, traditional measures based on religion or language are also simultaneously de-

[1]We do not address in this chapter the second part of Collier's argument. While the first conjecture is already to some extent contentious, the second part of the global argument is even more so. Research in this subject must overcome critical technical issues regarding migrants' self-selection. For example, Docquier et al. (2020) show that those who intend to emigrate display different attitudes and values when compared to the overall population in the origin country. Moreover, migrants tend to select countries that display high productivity translated into wages and GDP growth. Therefore, estimating the effect of the cultural diversity in the destination societies is not a simple task. Nevertheless, some studies have found an overall positive effect of diversity in economic growth after taking into account potential endogeneity. See Docquier et al. (2019) for a comprehensive literature review.

termined by migration patterns. For instance, the linguistic distance between the US and Mexico has been reduced with the extensive immigration of Mexicans to the US. Transfer of norms or adoption of new ones can influence religious practice, both in origin and destination countries, influencing the cultural distance itself. To deal with such an issue, we proxy cultural distance with ancestral distance or genetic distance, taking benefit of the recent work of Spolaore and Wacziarg (2009) and Spolaore and Wacziarg (2018). Geneticists use the frequencies of alleles of neutral genes to study genetic relatedness between populations. Similar to genes, cultural values are also transmitted from generation to generation (Richerson and Boyd, 2008; Bisin and Verdier, 2001). The positive and significant correlation between ancestral distance and cultural distance has been confirmed by Desmet et al. (2011) and Spolaore and Wacziarg (2016). Moreover, ancestral distance is an important factor that explains the development gap and the transmission of technological knowledge across countries (Spolaore and Wacziarg, 2009). We relate these measures with data on migration flows and stocks that are the central ingredients of the econometric specification of the network effect.

In order to confirm whether the network effect varies with the cultural distance between the origin country of the migrant and the destination, we estimate the interaction between these determinants. For this task, we estimate a gravity model to explain international migration inflows using the Poisson Pseudo Maximum Likelihood (PPML) estimator elaborated by Santos Silva & Tenreyro (2006). We find empirical evidence of a moderate interaction effect between ancestral distance and migrant networks. In summary, the elasticity effect of migrant networks is 0.8 when the migrants come from the farthest ancestral distant country with respect to the destination. In contrast, the elasticity effect of migration networks is 0.6 when the ancestral distance between the destination and origin countries is zero. Nevertheless, while it is statistically significant, the effect remains very moderate and of second order with respect to other factors such as migration policies.

Addressing the omitted variable bias in this type of estimation is a key element in understanding ancestral distance's role in the network effect. To account for the existence of potential omitted variable bias, we employ the use of dyadic fixed effects, as well as the instrumental variable approach. After controlling for omitted confounders, we found that the impact of ancestral distance on the network effect is statistically significant but small in magnitude. In addition, we test the robustness of the results to alternative measures of ancestral distance and potential measurement error in migration stocks.

The paper is organized as follows. Section 2 provides a selective review of the literature connected to our paper. In particular, we review some work on the network effect in international migration as well as the literature studying culture-related distance as a determinant of immigration flows. Section 3 sets up our empirical framework and presents the data used in our econometric analysis. Section 4 presents the main results and the robustness checks. Section 5 concludes.

2.2 Selected literature review

Our paper is related to two important strands of the economic literature on international migration. The first strand studies the effect of diaspora networks on international migration flows. This effect has been coined in the literature as the "network or the diaspora effect" in international migration. A rich body of literature has shown the importance of migrant networks to explain the size and the direction of international migration flows. While some studies focus on particular corridors, such as the Mexican migration to the US (Massey and Espinosa, 1997; Winters et al., 2001; Munshi, 2003; Garip and Asad, 2016; Mahajan and

Yang, 2017; McKenzie and Rapoport, 2010); others have analyzed the effect of migration networks in cross-country studies (Beine et al., 2011, 2015; Docquier et al., 2014; Bertoli and Fernández-Huertas-Moraga, 2012, 2015).

The majority of cross-country studies have found a positive and significant relationship between migrant networks and subsequent migratory flows. Migration networks are proxied in these studies using the stock of migrants from the origin country in each destination. For instance, Pedersen et al. (2008) have used yearly migration inflows to study the determinants of international migration. The authors gathered data on bilateral migration flows and stocks from statistical bureaus for 27 selected OECD destination countries during the period 1989-2000. Moreover, Beine et al. (2011, 2015) and Beine (2016) use the variations in stocks to proxy migration flows to study the effect of networks on their magnitude as well on the skill selection in migratory flows. The estimated elasticities show remarkable convergence in terms of point estimates across studies. These elasticities range between 0.45 and 0.6 for the skilled immigrants and between 0.6 and 0.8 for the unskilled ones. Also, Pedersen et al. (2008) estimated an elasticity of 0.59, which is fully consistent within these ranges.[2]

The observed network effects can be rationalized along with two different economic effects. The first one is the so-called assimilation effect. These networks consist of a set of interpersonal ties (kinship, friends, and shared communities) that affect the migration pro-

[2]While this positive relationship is consistent across studies, few of them have accounted for the endogeneity problem that arises in the estimation of network effects as argued by Manski (1993) and Munshi (2016). Previous migration networks and subsequent migrant flows are both determined by the same underlining factors, with many of them unobserved by the researcher. The endogeneity problem caused by the omission of unobserved variables has been dealt with in the literature using the instrumental variable approach. For example, Beine et al. (2011) use as instrumental variables the guest worker agreements in the 1960s and the 1970s, and a variable capturing the unobserved diaspora in the 1960s through a combination of push factors at origin, size, and openness to migration at the destination, and between the origin and destination country. Similarly, Bertoli and Fernández-Huertas-Moraga (2012) used migration stocks in 1960 as an instrumental variable for migration stocks in 1990. Both papers have found that the existence of migration networks within a country does, in fact, casually explain the future flows of migrants from the same origin.

cess by diminishing movement costs and risks. For example, migrants who settled earlier in the destination country may act as a provider of information, capital, and other resources to potential newcomers. At the destination, they can supply direct assistance to recent migrants in terms of housing, food, transport, credit; or serve in the adaptation process by providing psychological and social support.[3] Related to that, it has also been shown that the size of the network at the destination is related to higher employment probability and better-paid jobs at destination (Munshi, 2003).

The second crucial economic channel driving the network effect is family reunification. In most countries, kinship relationships allow to overcome the legal hurdles set by immigration policies. Family migration is an important alternative channel to economic migration, especially for potential migrants who do not fall within the visa policies. Unsurprisingly, the legal channel turns out to be more prevalent for unskilled immigrants, both in terms of magnitude and in terms of the share of the global network effect. Using immigration data in the US at the metropolitan level, Beine et al. (2015) find that the legal channel explains about 25% of the network effect for skilled migrants. At the same time, it can represent up to 50% for unskilled migrants coming from developing countries.

The second strand of literature focuses in the role of culture and language on migration decisions. The differences in culture and language between destination and origin countries might hinder the migrants' capacity to transfer their skills and abilities while decreasing their expected income (Dustmann and Glitz, 2011). If migrants internalize this effect, it might lead them to choose destinations culturally closer to their own in order to mitigate these disadvantages. The above has been argued by studies such as Adserà and Pytliková (2015).

[3]Comola and Mendola (2015) provides evidence of two specific assimilation effects. Giulietti et al. (2018) provide evidence on weak and strong ties. Bertoli and Ruyssen (2018) confirm the existence of strong assimilation effects by showing the strong effect of personal foreign connection on intended migration in the Gallup World Poll survey.

The authors show how emigration rates are higher among countries that share similarities in their languages. In other words, country pairs with higher linguistic distances will have lower migration flows between each other. Other studies have used religious proximity as a proxy for cultural distance. For instance, Belot and Ederveen (2012) use religious distance and measures of cultural distance from the World Values Survey to argue that cultural differences have a negative effect on international migration flows across 22 OECD countries.

While differences in culture might create barriers to migration, the existence of a large community of established migrants with a similar language and culture at the destination might mitigate the adaptation costs associated with migration, encouraging others from the same origin to move abroad. It might be that the mitigation effect of the migrant network is stronger at high levels of cultural distance between the country of origin and destination. Therefore, both cultural distance and the network of migrants explain migration flows jointly. Collier (2013) conjectures that the network effect is higher for migrants coming from more culturally distant countries. In subsequent work, Collier and Hoeffler (2018) provide empirical support to this hypothesis. While this hypothesis is an interesting intellectual and academic curiosity, if this is confirmed, it might also raise important questions and concerns for destination countries. For instance, large diasporas might encourage migrants to remain enclosed within their ethnic enclaves at the destination while keeping solid connections with their co-national at the origin. The above could hinder migrants' assimilation process in the host countries, especially for those coming from more distant cultures. Moreover, there could be unintended consequences for the host countries' institutions due to the presence of migrants from culturally far origins. The study of these potential consequences goes beyond the scope of this paper.

2.3 Data and descriptive statistics

2.3.1 Long-term cultural distance proxied by ancestral distance

In order to study how the network effect in international migration varies with cultural distance, we use the information on ancestral distance as a proxy of long-term cultural distance. The use of this variable relies on the fact that after the prehistoric movement of Homo sapiens out of Africa, human populations have dispersed spatially over the earth. The above forced human tribes to diverge genetically and culturally over time and across space. During that process, random gene mutations were likely to occur. These mutations in genes are transmitted from parent to child. Similarly, cultural attitudes are also transmitted from generation to generation through imitation, cultivation, instruction, and other types of social transmission (Richerson and Boyd, 2008; Bisin and Verdier, 2001). Hence, a large ancestral distance between two populations reflects not only how two populations have separated from the same *Homo* species but also how they have diverged in terms of intergenerationally transmitted attributes such as cultural values, norms, habits, and beliefs.

The correlation between ancestral distance and cultural distance has been studied by Desmet et al. (2011) and Spolaore and Wacziarg (2016). Desmet et al. (2011), compare genetic distance with cultural distances measured from the World Value Survey (WVS). The authors find a positive and significant correlation between genetic and cultural distance after controlling for geographic and linguistic distance. Similarly, Spolaore and Wacziarg (2016) confirms that ancestral distance measures are positively correlated with linguistic, religious distance and the diversity in values and norms across countries expressed in the WVS. The authors emphasize that ancestral distance is an aggregate measure of the different dimensions of cultural distance. Nevertheless, it should be that ancestral distance describes a long-term cultural distance and does not capture current changes in cultural values, norms, or habits.

To measure ancestral distance, geneticists calculate the variation in allele frequency between populations. According to the Encyclopædia Britannica, an allele is a variation of a particular gene. Every gene resides in a specific *loci* or location in the chromosome in two copies. Each copy corresponds to information inherited from each parent. While most genes are the same within a species, some of them have different forms. These are known as alleles. A gene can have different alleles within a species due to small mutations that occur as a consequence of natural selection or random drift. Geneticists use the frequencies of alleles of neutral genes that change randomly over time to study genetic relatedness between populations. If the frequency of all alleles is identical between two communities, the measure of ancestral distance will be equal to zero, meaning that the two populations are genealogically related (Cavalli-Sforza et al., 1994). On the contrary, the larger the ancestral distance is, the more time has passed since the two populations shared a communal ancestor.

Ancestral distance measures have been used before in economic literature. For example, Spolaore and Wacziarg (2009) study the effect of ancestral distance on the diffusion of economic development using genetic data calculated by Cavalli-Sforza et al. (1994). Cavalli-Sforza et al. (1994) employed classic genetic markers from 42 populations worldwide and 120 *loci*. Recently, Spolaore and Wacziarg (2018) updated their results using new data on human genome microsatellite variation to measure the relatedness between societies. The authors match genetic data by ethnicities from Pemberton et al. (2013) to countries using ethnic composition data from Alesina et al. (2003). The new data collected by Pemberton et al. (2013) improves on the original as it provides comprehensive coverage of the 267 populations from 645 microsatellites *loci*. This version allows for a better match between population and ethnic groups, especially those located in Asia and Africa (Spolaore and Wacziarg, 2018).

Since most countries are typically inhabited by different ethnic groups, Spolaore and Wacziarg (2018) calculated two indicators of genetic distance. The first one, called "Ancestral distance, dominant", is the distance between the ethnic groups with the largest share in the population of each country. The second one, coined "Ancestral distance, weighted", is a weighted measure of genetic distance that includes all ethnic populations within a country. These two measures of genetic distance show a correlation coefficient of 0.91. However, the values differ for those countries with large genetic admixtures. We make use of the new weighted ancestral distance measure in our benchmark results since it captures the ethnic composition within origin countries. This could be important for origins with large ethnic diversity. Moreover, we consider that the quality of information is superior in the new measure since it not only covers more genetic populations but also uses more genetic information, reducing measurement error. Nevertheless, in the robustness checks, we also employ the other alternative measures of ancestral distance. The countries with the most significant average ancestral distance (weighted) with respect to our 32 destinations are Solomon Island, Tonga, and Vanuatu (see Figure 4 in Appendix A).

2.3.2 International migration inflows

Our econometric analysis relies on the bilateral immigration flows database compiled by the International Migration Institute (DEMIG, 2015a). We use international migration flow data instead of changes in migrant stocks over time since the latter correspond to net migration. The database contains yearly data on international migration inflows and outflows from 236 origins to 33 destination countries over the period 1946–2011[4]. In addition to typically used OECD destinations, it includes other countries such as South Africa, Brazil, Mexico, and

[4]We exclude Czechoslovakia from our analysis.

Uruguay to allow for a larger variability in the genetic composition of the sample. DEMIG (2015a) reports information gathered from primary national sources without making major adjustments, alterations, or calculations to fill in missing values. The data combines different sources: population registries, border statistics, and permits (residence/work) published in annual statistical, demographic, and migration yearbooks on behalf of national statistics offices.

The public version of the data, however, does not report information for some destinations countries such as Norway, Canada, Australia, the Netherlands, and Italy, at least for some sub-periods of time. We, therefore, supplement the public DEMIG database with migration flows collected from these countries directly. Appendix A reports the sources of the data for the five countries mentioned above. A further issue is that DEMIG (2015a) database does not contain a full breakdown of all origin countries for specific destinations such as Poland, Slovenia, Luxembourg, and the United Kingdom. To solve this issue, we opted to use the international migration bilateral flows dataset from the UN international flows database (United Nations and Social Affairs, 2015) for these four destinations. The two data sources (DEMIG and UN) were gathered from similar sources. Nevertheless, we compared the available pairs between the two datasets. We calculated the pairwise correlation coefficient of 0.97 between the two datasets when the exact migration definition was available (See Figure 2.2).

In order to construct our final database, we proceed with the four following steps. First, we select a migration definition depending on the number of origins and years available for each destination. Three potential criteria are used to define migrants: by country of birth, by country of the last residence, or by country of citizenship. Table 2.8 in Appendix A2 shows in detail the migration definition used for each destination country. Second, for some countries whose constitution has changed over time, we collapsed the data into single entities to ensure comparability. This concerns East and West Germany. Third, we fill up the missing years within each decade for some corridors. To that aim, we opted to sum up the migration flows over ten years and multiply this sum by $10/i$, where i is the number of non-missing years within the decade. Fourth, many destination countries often do not report specific origins, especially those from which they do not receive migrants. In that case, destinations that report large numbers missing values do not report zero values. This is the case for Brazil, Czech Republic, Finland, France, Israel, Portugal, Spain, and Uruguay. Therefore, we fill up the empty cells for these non-reported corridors with zeroes if we observe zero migration in the Global Bilateral Migration Database from Ozden et al. (2011). In the case of a non-zero value in Ozden et al. (2011), we treat the flow as a missing observation. From the overall sample of 17394 observations used in the estimations, 23% correspond to imputed values. Our final sample constitutes migration inflows data from 1960 to 2009 for 175 sending countries to 32 destination countries. The list of origin countries is presented in Appendix A.

2.3.3 Other data

The size of the existing migration diaspora is measured as the stock of migrants from origin i to destination j at the beginning of the decade published by World Bank in the Global Bilateral Migration Database from Ozden et al. (2011). We also control for colonial links, contiguity, geographical and common language from Head and Mayer (2014).

Table 2.1: Descriptive Statistics of migrations flows, stocks and ancestral distance

	Mean	Std. Dev.	Min.	Max.
Migration flows	7,949.502	60485.27	0	3,374,712
Migration stocks	1,4262.37	114765.1	0	9,367,910
log Geodesic distance (Km^2)	8.647	0.874	4.087	9.884
Contiguity	0.024	0.155	0	1
Colonial link	0.028	0.167	0	1
Common language	0.126	0.332	0	1
Ancestral distance (weighted)	0.029	0.018	0	0.085
Ancestral distance (dominant)	0.030	0.022	0	0.097
Ancestral distance (weighted, old)	0.090	0.064	0	0.311
Ancestral distance (dominant, old)	0.091	0.075	0	0 .337
Observations	17394			

Notes: Old ancestral distance variables refers to data obtained from Cavalli-Sforza et al. (1994).

2.4 Empirical Strategy

2.4.1 Econometric specification

We first present our benchmark specification. This specification is consistent with the equilibrium derived from a Random Utility Maximisation model in which prospective migrants maximize their utility across all possible locations in the world, including home (see for instance Grogger and Hanson (2011),Beine et al. (2011, 2015), among others). The benchmark specification that we estimate is given by:

$$ln(N_{ij,t}) = \beta_1 ln(1 + M_{ij,t}) + \beta_2 AD_{ij} + \beta_3 AD_{ij} * ln(1 + M_{ij,t}) + \gamma X_{ij} + \alpha_{j,t} + \alpha_{i,t} + \epsilon_{ijt} \quad (2.1)$$

where $(N_{ij,t})$ is the migration flow over the decade t from country i to country j. The total size of the diaspora at the beginning of the decade is captured by the M_{ijt}. Following Beine et al. (2011) and Bertoli and Fernández-Huertas-Moraga (2012), we assume a logarithmic form for the size of the diaspora and add one to avoid losing observations due to the functional form when migration stocks are zero. The variable AD_{ij} corresponds to the measure of ancestral distance from Spolaore and Wacziarg (2009, 2018). We have standardized AD_{ij} (subtracted the mean and divided by the standard deviation) for comparison purposes between the different measures of ancestral distance.[5] We include a set of observable bilateral variables X_{ij} to control for geographical distance, shared borders, former colony dummy, and common language. The variables are resumed in Table 1.

To account for multilateral resistance to migration (Bertoli and Fernández-Huertas-Moraga, 2013), we include a set of origin-time and destination-time fixed effects. Moreover, destination-time fixed effects allow us to control for factors that change over time and explain the international migration flows to a specific destination, such as changes in migration policies. It also captures differences in migration definitions reported by each country and policy interventions such as amnesties or regularisation. Similarly, origin-time fixed effects capture specific factors at origin that influence international migration, such as macroeconomic shocks, changes in the average wage level at origin country, or wars. As explained by Beine and Parsons (2015), $\alpha_{i,t}$ also captures the percentage of people at the origin who decide to not migrate from their origin countries.

2.4.2 Econometric issues

Identifying the network effect on migration flows is often challenging due to the potential endogeneity of the diaspora. The coefficients β_1 and β_3 might be biased due to the omission of other unobserved factors that are potentially correlated with both the error term and the

[5] Notice that we have included AD instead of $ln(AD)$. The main reason is that a large number of country pairs share zero ancestral distance. Therefore, the inclusion of $ln(AD)$ will lead to a significant number of missing values. Unlike migration stocks, ancestral distance ranges from 0 to 0.08. Nevertheless, results are consistent when using $ln(AD + 1)$ (See Appendix A).

network. Furthermore, Giuliano et al. (2013) argue that the simple measure of geographical distance between countries is not sufficient to capture other geographical conditions that might hinder or facilitate gene flow and admixture between ethnic populations, such as mountain chains or access to the same rivers or seas. These factors might be omitted in equation (2.1). We first account for time-invariant omitted variables, by extending equation (2.1) to include of dyadic fixed effects ($\alpha_{i,j}$) as expressed in equation 2.2. These fixed effects will capture other dimensions of cultural proximity that are time-invariant such as similar language, ethnicity, or religion.[6] We estimate the following equation:

$$ln(N_{ij,t}) = \beta_1 ln(1 + M_{ij,t}) + \beta_3 AD_{ij} * ln(1 + M_{ij,t}) + \alpha_{j,t} + \alpha_{i,t} + \alpha_{i,j} + \epsilon_{ijt} \tag{2.2}$$

Moreover, estimating equations (2.1) and equation (2.2) using Ordinary Least Squares (OLS) might lead to biased and inconsistent estimates of β_1 and β_3 in the presence of heteroskedasticity and a high number of zeros in the dependent variable (Santos Silva and Tenreyro, 2006). International migration is often equal to zero in some corridors as a result of the existence of highly restrictive migration policies or very low gains from migration for these corridors. As expected, our measure of migration flows is equal to zero for an important number of observations (31,9% of the observations). In that case, Santos Silva and Tenreyro (2006) suggest using the bilateral flows as the dependent variable $N_{ij,t}$ and estimate the model using Poisson Pseudo-Maximum Likelihood (PPML). Using Monte Carlo simulations, the authors compare the performance of OLS and PPML estimators and show that the PPML estimator is unbiased and consistent under the existence of a large percentage of

[6] With the inclusion of a_{ij} fixed effects, our source of variation is over time within the ij unit. Moreover, the errors are no longer independent and identically distributed (i.i.d.) and have autocorrelation within the unit of analysis. In order to take into account the correlation of the error terms over time within country pairs, the standard errors are clustered at the dyadic pair level.

zeros and heteroskedasticity. For this reason, we estimate equation (2.1) and (2.2) using the PPML estimator.

2.5 Results

2.5.1 Benchmark results

Table 2.2 presents the benchmark estimation results for the determinants of international migration flows, including weighted ancestral distance. The first column shows a replica of the results presented by Beine (2016) using international migration inflows as a dependent variable instead of the difference in stocks between decades. These alternative information sources lead to similar estimates of the network effect on international migration flows without considering other omitted bilateral determinants. In line with previous studies such as Pedersen et al. (2008) and Beine et al. (2011), we estimate a network elasticity that ranges from 0.58 to 0.69 depending on the specification. The coefficients of the other control variables are highly significant and in the expected direction, with the exception of contiguity, which is not significant. As expected, we find that an increase in the geographical distance between the country of origin and destination decreases migration inflows. Moreover, we find that sharing a similar language boosts migration flows by 6 percent.

Furthermore, column (2) in table 2.2 shows the results when ancestral distance is added to the model. Similar to Adserà and Pytliková (2015), the coefficient of ancestral distance is small and not significant. However, once the interaction between migrant networks and weighted ancestral distance is accounted for, we find that ancestral distance has a negative and significant effect on migration flows depending on the level of the diaspora of migrants from the same origin in a particular destination (See column 3). Given that we are estimating a non-linear model, we present in the appendix section the calculated marginal effect of

Table 2.2: PPML estimations of migration flows between 1960 - 2009 using ait, ajt and aij fixed effects

	(1)	(2)	(3)	(4)
log (stocks+1)	0.586***	0.583***	0.618***	0.698***
	(0.0209)	(0.0205)	(0.0210)	(0.0233)
Ancestral dist. (weighted)		-0.172	-0.740***	
		(0.151)	(0.171)	
log (stocks+1) x Ancestral dist.			0.0664***	0.0364**
			(0.0156)	(0.0136)
log (geodesic distance)	-0.308***	-0.303***	-0.293***	
	(0.0546)	(0.0564)	(0.0560)	
Contiguity	-0.270	-0.270	-0.230	
	(0.164)	(0.164)	(0.166)	
Colonial link	0.206*	0.213*	0.233**	
	(0.113)	(0.112)	(0.111)	
Common language	0.677***	0.678***	0.679***	
	(0.0904)	(0.0903)	(0.0902)	
Observations	17394	17394	17394	18229
R^2	0.922	0.922	0.923	0.903
Dyadic FE ($\alpha_{i,j}$)	No	No	No	Yes
Origin-time FE ($\alpha_{i,t}$)	Yes	Yes	Yes	Yes
Destination-time FE ($\alpha_{j,t}$)	Yes	Yes	Yes	Yes

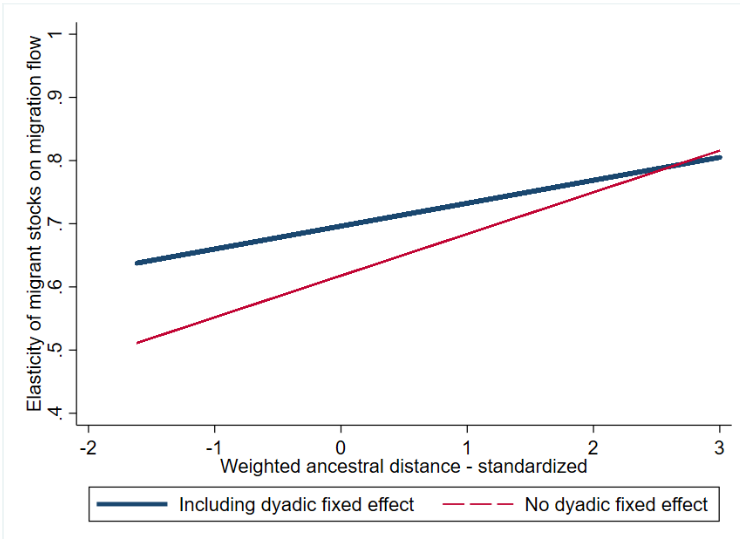
Notes: Ancestral dist. (weighted) has been re-scaled to have a mean of zero and a
The standard errors are presented in parentheses clustered at the country pairs level.
bilateral pair level. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

ancestral distance in Figure 2.3. The above shows that the long-term cultural distance is negatively associated with international migration flows, but its effect depends on the magnitude of the existing stock of migrants.

In addition, column 3 also displays the coefficient of the interaction between migration stocks and ancestral distance. We find a positive interaction effect between ancestral distance and migration networks. This positive interaction effect shows that as soon as ancestral distance increases, the effect of the previous diaspora from the same country at the destination is associated with larger international migration flows to this destination (See column 3). Notice that throughout columns (1) to (3), we have included the classical two types of fixed effects (i.e., origin country x time fixed effects and destination-country times fixed effects). Moreover, column (4) in table 2.2, shows the results when we include the full set of dyadic fixed effects into the model.[7] Figure 2.1 presents the estimated marginal effects of the stock of migrants at different levels of ancestral distance graphically. The elasticity of migration stocks on migration flows ranges from 0.6 to 0.8, depending on the level of ancestral distance. We find that when we control for unobserved time-invariant factors, the coefficient of the interaction between ancestral distance and the network reduces by half, meaning that the elasticity of migration stocks on flows is not as dependent on ancestral distance as it was considered by Collier and Hoeffler (2018). Therefore, while migrant networks are more relevant for migrants who face larger cultural differences with respect to the host society, this effect is small when many omitted factors are included.

[7] The dyadic fixed effects capture the effect of time-invariant variables on migration inflows; therefore, we cannot estimate the effect of ancestral distance. Nevertheless, we can still identify the network effect and its interaction with ancestral distance.

Figure 2.1: Marginal effects of migration diasporas on inflows at different levels of genetic distance from column (3) and (4) in Table 2.2.



2.5.2 Robustness checks

2.5.2 Omitted variable bias

The inclusion of time-invariant dyadic fixed effects might be insufficient to account for the underlying omitted variable bias related to the presence of the network in equation (2.1). For instance, if unobserved determinants such as time-variant cultural distance influence both the diaspora and migration inflows, the inclusion of dyadic fixed effects will not eliminate this bias. As a result, it might be desirable to estimate equation (2.2) with an instrumental variable approach. One caveat resulting from the use of exponential models such as PPML within the instrumental variable model is the difficulty of the inclusion of high-dimensional fixed effects. Despite the development of a Generalized Method of Moments (GMM) framework for count data models with endogenous regressors by (Santos Silva and Windmeijer, 1997), we encounter convergence issues when using Poisson-IV and including more than one layer of fixed effects (see Tenreyro, 2007). For this reason, we rely on the traditional Two Step-Least Square (2-SLS), for which we transform the dependent variable and estimate equation (1) using $ln(N_{ij}+1)$ as the dependent variable. The former allows us to keep the zero migration flows within the estimated sample.

Table 2.3 presents the instrumental variable results of the determinants of migration flows. We use a 2-SLS model with the predicted migration stocks (Mo_ij) and its interaction with ancestral distance as two instrumental variables. We predict the stock of migrants for each decade using the share of immigrants from each origin to a J destination in the decade 1960 and the total immigration to destination country J in each subsequent decade.[8]

Since we use the data from 1960 to construct the instrumental variables, the results displayed in Table 2.3 excluded this decade from the sample. For comparison reasons, we show in column (1) the estimation of equation (2.1) using data from 1970 until 2009 under PPML. Moreover, we also present the estimation of equation (2.1) using scaled OLS in column (2) in order to compare these results with the 2-SLS results presented in column (3). Columns (4) and (5) display the first-stage regression results. The instrumental variables pass the F-stat test on excluded instruments, meaning that these are good predictors of the endogenous variables. The addition of a constant to deal with the zero counts introduces a bias in the model (Winkelmann, 2008, pg.66); however, the results of the IV estimation lead to similar conclusions as Table 2.2. In particular, the coefficient of the interaction effect is quantitatively speaking, in line with the estimated coefficient using the full set of dyadic, destination-time, and origin-time fixed effects presented in Table 2.2. Column (3) shows an interaction coefficient equal to 0.028 compared to a coefficient equal to 0.036 displayed in

[8] The use of these type of "shift-share" instrumental variables have been criticized by Jaeger et al. (2018) due to the presence of a high degree of serial correlation in the country-of-origin distribution of immigrants. However, it has been difficult - not to say impossible - task to find an exogenous variable with an a_ijt dimension that is correlated with the past diaspora and uncorrelated with the error term.

column (4) in Table 2.2.

Moreover, we extend equation (2.2) with the inclusion of bilateral travel visas as a proxy of visa restrictions. Essentially, this type of visa acts as a barrier to control for the entry of particular migrants, for example, asylum seekers (Czaika et al., 2018) or temporal workers. The above partially control for the omitted time-varying dyadic factors contained in the error term. If changes in cultural distance correlate with changes in migration policies, then the inclusion of the former control could capture some of the omitted determinants of migration flows. Table 2.4 presents the results with this additional control variable. Using DEMIG's annual tourist visa data (DEMIG, 2015b), we construct a variable that is equal to one if a national from the country of origin needed a tourist visa to enter into a particular destination for more than six years within a decade. The DEMIG Visa data is available from 1973 until 2013; therefore, we complement this database using Mau et al. (2015) for the year 1969. The scope of this variable is limited in terms of countries and time. The above means that the number of missing values increases. For this reason, columns (1) and (2) replicates the results from columns (3) and (4) in Table 2.2 using the sample of non-missing observations for the variable visa restrictions.

Table 2.3: Estimations of Migration Flows with instrumental variables using 2SLS

	(1)	(2)	(3)	(4)	(5)
	PPML	OLS	2SLS	First Stage	
	Inflows	log (Inflow+1)	log (Inflow+1)	log diasp.	log Dias. x AD
log (stocks+1)	0.625***	0.428***	0.504***		
	(0.0208)	(0.0120)	(0.0200)		
log(stocks+1) x Ancest. dist.	0.0597***	0.000400	0.0285***		
	(0.0146)	(0.00789)	(0.00947)		
Ancestral dist. (weighted)	-0.707***	0.0862	-0.0279	-0.370***	1.552***
	(0.161)	(0.0644)	(0.0679)	(0.0763)	(0.0846)
log geodesic dist.	-0.302***	-0.586***	-0.525***	-0.353***	-0.198***
	(0.0511)	(0.0409)	(0.0399)	(0.0352)	(0.0463)
Contiguity	-0.315**	-0.246	-0.237	0.556***	-0.899***
	(0.127)	(0.193)	(0.187)	(0.135)	(0.154)
Colonial link	0.313***	1.358***	1.217***	0.870***	0.00833
	(0.0968)	(0.180)	(0.175)	(0.153)	(0.140)
Common language	0.675***	0.614***	0.486***	0.407***	0.280***
	(0.0818)	(0.0817)	(0.0809)	(0.0677)	(0.0726)
lnMoj				0.593***	0.0107
				(0.0111)	(0.0128)
lnMojxAD				0.0328***	0.806***
				(0.00643)	(0.00870)
Observations	15637	15637	15637	15637	15637
R^2		0.817	0.816		
Partial R^2				0.357	0.673
First stage F-stat				1439.336	4309.531
Dyadic FE ($\alpha_{i,j}$)	No	No	No	No	No
Origin-time FE ($\alpha_{i,t}$)	Yes	Yes	Yes	Yes	Yes
Destination-time FE ($\alpha_{j,t}$)	Yes	Yes	Yes	Yes	Yes

Notes: Moij are the predicted stocks using 1960 shares. From (1) to column (5), the 1960 are excluded from the sample The Angrist and Pischke (2009) R-squared is reported for the first stage. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
The standard errors are presented in parentheses clustered at the country pairs level at the bilateral pair level.

The third column of table 2.4 shows the results when visa restrictions are added to the estimations. Moreover, column (4) shows the results when dyadic fixed effects are included as well. Overall, the estimations are consistent with the inclusion of this variable. Importantly, the interaction effect remains positive and significant; however, the estimated coefficient of the interaction is slightly higher when we control for visa restrictions. The coefficient for the interaction effects is 0.038 without including visa restrictions as a control variable, and it is 0.042 after controlling for traveling visa restrictions. Nevertheless, the general conclusion deduced from Table 2.2 remains.

Moreover, in column (5) of table 2.4, we include an interaction between migrant networks and visa restrictions. We found that while tourist visa restrictions have a negative effect on international migration flows, the presence of the stock of migrants of the same country mitigates this adverse effect. Migrant networks exert a stronger impact on migration flows in the presence of visa restrictions, as shown in the interaction effect. The positive interaction effect between visa restrictions and the network is consistent with the legal channel of the network effect (Beine et al., 2015). It fully supports the hypothesis that migrations networks act as a conduit to overcome legal barriers. This interaction is, quantitatively speaking, larger when compared to the interaction between ancestral distance and migration stocks. Controlling for travel visa restriction and its interaction with the stock of migrants does not change the coefficient of the interaction between cultural distance and existing migrant stocks.

Table 2.4: PPML estimations of Migration Flows including visa restrictions

	(1)	(2)	(3)	(4)	(5)
log (stocks+1)	0.619***	0.697***	0.621***	0.700***	0.693***
	(0.0157)	(0.0233)	(0.0157)	(0.0214)	(0.0215)
Ancestral dist. (weighted)	-0.730***		-0.756***		
	(0.121)		(0.122)		
log(stocks+1) x Ancest. dist.	0.0640***	0.0381***	0.0666***	0.0427***	0.0333***
	(0.0111)	(0.0136)	(0.0110)	(0.0110)	(0.0127)
log geodesic dist	-0.288***		-0.259***		
	(0.0354)		(0.0370)		
Contiguity	-0.213*		-0.184		
	(0.123)		(0.122)		
Colonial link	0.232***		0.196**		
	(0.0842)		(0.0856)		
Common language	0.666***		0.671***		
	(0.0657)		(0.0656)		
Visa restrictions			-0.319***	-1.107***	-0.962***
			(0.0694)	(0.149)	(0.150)
log(stocks+1) x Visa restrictions					0.322***
					(0.105)
Observations	16009	16459	16009	16459	16317
R^2	0.924	0.903	0.924	0.916	0.916
Dyadic FE ($\alpha_{i,j}$)	No	Yes	No	Yes	Yes
Origin-time FE ($\alpha_{i,t}$)	Yes	Yes	Yes	Yes	Yes
Destination-time FE ($\alpha_{j,t}$)	Yes	Yes	Yes	Yes	Yes

The standard errors are presented in parentheses clustered at the country pairs level. Visa restrictions is a dummy variable equal to 1 when there have been restrictions to travel during the majority of the decade. It is zero otherwise. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.5.2 Sample-specific estimations

Table 2.5 shows the results of the effect of the network and the interaction with ancestral distance when we split the sample into corridors by high or low skill selection. We use data on bilateral migration stocks by gender and education in 2000 from Artuc et al. (2015). Since the data is only available for a few decades, we restrict our estimations to the decade of 2000. Column (1) shows the results of equation (2.1) for all the dyads with non-missing information. Columns (2) and (3) show the results for corridors where migrants are mostly low-skilled and high-skilled, respectively. As expected, the results reveals that the network effect on migration inflows is higher for those corridors with a higher incidence of low-skilled migrants than for corridors with predominantly high-skilled migrants. The interaction between the stock of migrants and ancestral distance is slightly higher for those corridors with abundantly low-skilled migration. However, the size of these effects remains low.

Moreover, dividing the sample according to above or below the median skill ratio shows that cultural distance is a relevant determinant of migration flows for corridors with a predominance of unskilled migrants. The larger the ancestral distance is, there will be less migration coming from these countries. This result is consistent with Krieger et al. (2018), whose article shows that at higher levels of ancestral distance, low-skilled migrants are strongly discouraged from migrating while high-skilled migrants are neither encouraged nor discouraged by the ancestral distance.

Moreover, we evaluate if the results are driven by the quality of the data. Migration stocks were imputed by Ozden et al. (2011) in cases when countries went through a separation process, such as the former USSR, Czechoslovakia, and Yugoslavia. In addition, migration stocks were extrapolated for countries that had missing census data and when some origin countries were aggregated in broad categories such as "Other Africa". Columns (1) and

(3) in table 2.6 excludes those destinations countries with the missing census in a particular decade and those countries that went through a separations process. We exclude, in addition, the origin countries that were potentially collapsed in broad categories and extrapolated by Ozden et al. (2011). The results without all potential imputations in the stocks are presented in Columns (2) and (4) of table 2.6. Similar to the previous tables, table 2.6 uncovers a positive interaction term between ancestral distance and the network of migrants from the same origins that is still significant at the ten percent level. When compared to the results presented in 2.2, the estimated coefficient of the interaction is slightly smaller, but the coefficient of the stock of migrants is slightly bigger.

2.5.2 Alternatives measures of ancestral distances

Table 2.7 presents robustness checks to alternatives measures of ancestral distances. Columns (1) to (4) reports the estimates of equation (2.1) using the four ancestral distance indicators: the weighted and dominant measures calculated by Spolaore and Wacziarg (2009) and the same two measures calculated by Spolaore and Wacziarg (2018) using genome microsatellite variation. Similarly, Columns (5) to (8) reports the estimates of equation (2.2) using the four alternative measures and including dyadic fixed effects. The results show a small difference in the interaction coefficient between migration stocks and the two "new" measures of ancestral distance calculated in Spolaore and Wacziarg (2018). However, there are important differences in the interaction coefficient when we use instead the "older" measures calculated in Spolaore and Wacziarg (2009). In columns (3) and (4), the estimated interaction effect (0.086 and 0.079, respectively) are much larger than the estimated coefficients when using the "new" measures (0.0664 and 0.0669, respectively).

Moreover, the introduction of country-pair fixed effects, presented in columns (7) and

Table 2.5: PPML estimations of Migration Flows in 2000 by Skill Selection (HSij/Lsij)/(HSi/LSi)

	(1) Total	(2) Below median	(3) Above median
log (stocks+1)	0.630***	0.654***	0.495***
	(0.0255)	(0.0326)	(0.0426)
log(stocks+1) x Ancest. dist.	0.0495***	0.0547**	0.0488**
	(0.0163)	(0.0229)	(0.0235)
Ancestral dist. (weighted)	-0.946***	-0.923***	-0.108
	(0.193)	(0.268)	(0.235)
log geodesic dist.	-0.298***	-0.253***	-0.603***
	(0.0516)	(0.0501)	(0.103)
Contiguity	-0.185	-0.0700	-0.509
	(0.128)	(0.122)	(0.338)
Colonial link	0.444***	0.305**	1.059***
	(0.110)	(0.132)	(0.261)
Common language	0.618***	0.555***	0.537***
	(0.0843)	(0.107)	(0.113)
Observations	4574	2026	2535
R^2	0.904	0.936	0.936
Destination FE	Yes	Yes	Yes
Origin FE	Yes	Yes	Yes
Dyadic FE ($\alpha_{i,j}$)	No	No	No
Origin-time FE ($\alpha_{i,t}$)	No	No	No
Destination-time FE ($\alpha_{j,t}$)	No	No	No

Notes: The standard errors are presented in parentheses clustered at the country pairs level
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: PPML estimations of Migration Flows without imputed network levels

	(1)	(2)	(3)	(4)
log (stocks+1)	0.641***	0.641***	0.710***	0.709***
	(0.0225)	(0.0228)	(0.0264)	(0.0267)
log(stocks+1) x Ancest. dist.	0.0650***	0.0646***	0.0281*	0.0278*
	(0.0177)	(0.0179)	(0.0144)	(0.0145)
Ancestral dist. (weighted)	-0.837***	-0.839***		
	(0.187)	(0.189)		
log geodesic dist.	-0.336***	-0.337***		
	(0.0616)	(0.0617)		
Contiguity	-0.275	-0.275		
	(0.186)	(0.186)		
Colonial link	0.225*	0.225*		
	(0.122)	(0.122)		
Common language	0.662***	0.661***		
	(0.0916)	(0.0916)		
Observations	12445	12074	12697	12346
R^2	0.932	0.932	0.906	0.906
Dyadic FE ($\alpha_{i,j}$)	No	No	Yes	Yes
Origin-time FE ($\alpha_{i,t}$)	Yes	Yes	Yes	Yes
Destination-time FE ($\alpha_{j,t}$)	Yes	Yes	Yes	Yes

The standard errors are presented in parentheses clustered at the bilateral pair level. Columns (1) and (3) excludes destinations countries with missing census in a particular decade and those origins that went through a separation process such as the former USSR. Columns (2) and (4) show the results excluding also those origins that were collapsed originally in broad categories such as "other Africa"
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(8), reduces the interaction coefficient significantly to the point that the coefficient is no longer significant. One explanation is measurement error given the quality of the data used by Spolaore and Wacziarg (2009). The genetic data on 42 ethnic groups from Cavalli-Sforza et al. (1994) was matched to around 1,120 ethnic groups in Alesina et al. (2003). Many ethnic groups were paired with larger ones. For example, Afghan groups like "Balochi" and "Hazara" were categorized as "Iranians" as explained by Spolaore and Wacziarg (2018). The above introduces a significant measurement error in the independent variable. If the error in the measurement of AD_{ij} is unobserved and it becomes part of the error ϵ_{ijt}, then this type of measurement error can cause some bias in the estimated slopes. The greater the measurement error in the independent variable, the closer will the estimated slope approach zero instead of the correct value. In other words, it could lead to *attenuation bias*. Meaning that the estimated effect using the "old" measure will be attenuated as pointed by Krieger et al. (2018). We consider that the "new" ancestral distance indicators from Spolaore and Wacziarg (2018) are measured with less error than the "older" measures since the "new" measures use data from 267 populations from 645 microsatellites *loci*. This larger number of ethnic populations allows for a better-quality match with the ethnic groups of Alesina et al. (2003).

Table 2.7: PPML estimations of migration flows using different measures of genetic distance.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log (stocks+1)	0.618***	0.616***	0.624***	0.624***	0.698***	0.695***	0.680***	0.670***
	(0.0210)	(0.0211)	(0.0210)	(0.0210)	(0.0233)	(0.0222)	(0.0248)	(0.0228)
Ancestral distance (weighted)	-0.740***							
	(0.171)							
l(stocks+1). x Ances. dist. (weighted)	0.0664***				0.0364**			
	(0.0156)				(0.0136)			
Ancestral distance (dominant)		-0.429**						
		(0.206)						
l(stocks+1) x Ances. dist. (dominant)		0.0669***				0.0342***		
		(0.0133)				(0.0122)		
Ancestral distance (dominant, old)			-0.890***					
			(0.144)					
l(stocks+1) x Ances. dist. (dominant, old)			0.0861***				0.0172	
			(0.0136)				(0.0129)	
Ancestral distance (weighted, old)				-0.983***				
				(0.161)				
l(stocks+1) x Ances. dist. (weighted, old)				0.0797***				0.00810
				(0.0157)				(0.0133)
Observations	17394	17394	17394	17394	18229	18229	18229	18229
R^2	0.923	0.926	0.926	0.924	0.903	0.905	0.900	0.901
Dyadic FE ($\alpha_{i,j}$)	No	No	No	No	Yes	Yes	Yes	Yes
Origin-time FE ($\alpha_{i,t}$)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination-time FE ($\alpha_{j,t}$)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Regression includes bilateral control variables such as colonial links, contiguity, geographical distance and common language.
Ancestral distince measures were re-scaled to have a mean of zero and a standard deviation of one.
The standard errors are presented in parentheses clustered at the country pairs level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.6 Discussion and conclusion

Throughout this paper, we have studied the interaction between two effects that determines international migration flows: the networks effect and long-term cultural distance proxied by ancestral distance. While cultural distance might deter some individuals from moving to specific destinations, the presence of other co-national at the destination might help circumvent the adaptation costs associated with migration. The existence of a large community of co-nationals might encourage people to migrate to destinations where natives do not share similar values or norms of the potential migrant. Collier (2013) proposes this conjecture: the network effect is larger at high levels of cultural distance, suggesting that in a few decades, the flow of migrants coming from more culturally distant countries will increase dramatically. To empirically test this idea, this chapter uses annual inflow data of migrants into 32 destinations. It estimates the interaction effect of ancestral distance and the pre-existing migrant stocks on migration inflows.

Moreover, we account for potential problems related to the estimation of the network effect and perform numerous robustness checks. Identifying the network effect on migration flows is often challenging due to the potential endogeneity. One source of endogeneity is the omission of unobserved determinants of the migration flows at the dyadic level. To mitigate this bias, we employ dyadic fixed effects, include additional bilateral determinants such as visa restrictions, and apply an instrumental variable approach under 2SLS. Once these omitted determinants are controlled for, the estimated interaction effect reduces by half ranging from 0.02 to 0.04. Moreover, we estimate the network effect and its interaction with ancestral distance under different samples. First, we split the sample into corridors by high or low skill selection. Second, we exclude countries of destination for which the inflow data was imputed. Third, we exclude origin countries for which the inflow data was imputed as well. Finally, we replicate the results using alternative measures of ancestral distance.

Our preferred specification, using dyadic fixed effects and the weighted measure of ancestral distance using DNA microsatellites markers, shows an elasticity effect of migration networks on migration flows that oscillated between 0.6 and 0.8 depending on the level of ancestral distance. This means that when the population of destination and origin of the migrants share a communal ancestor and hence share cultural traits, the estimated elasticity is approximately 0.637. In addition, when the ancestral distance between the population of destination and origin is the largest in the sample (the case of South Africa and the Solomon Islands), the estimated elasticity is roughly 0.805. These results suggest that for more culturally distant countries, the diaspora effect fulfills a more important role in determining the immigration inflow than for countries that share a communal ancestral history. Nevertheless, the difference in the magnitude of network effect between these two types of country pairs is relatively small. Nowadays, other aspects such as migration policies or education have a more meaningful role in determining the composition of migrants in a particular destination.

2.7 Appendix A

2.7.1 Appendix A1. Correlation between UN and DEMIG migration flow

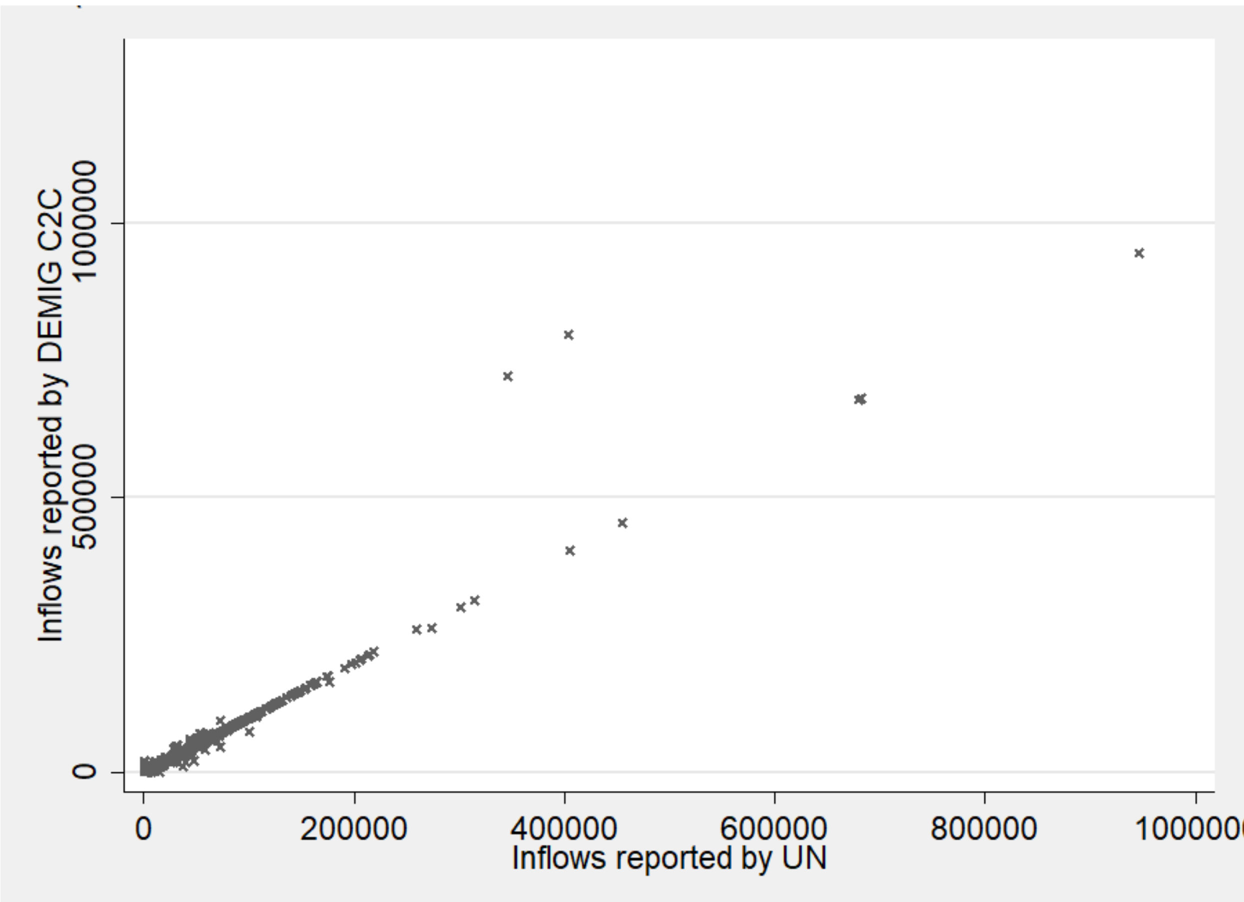


Figure 2.2: Comparing UN and DEMIG migration flow

2.7.2 Appendix A2. Definitions of international migrant by destination country

Table 2.8: Criteria to define migration by reporting country

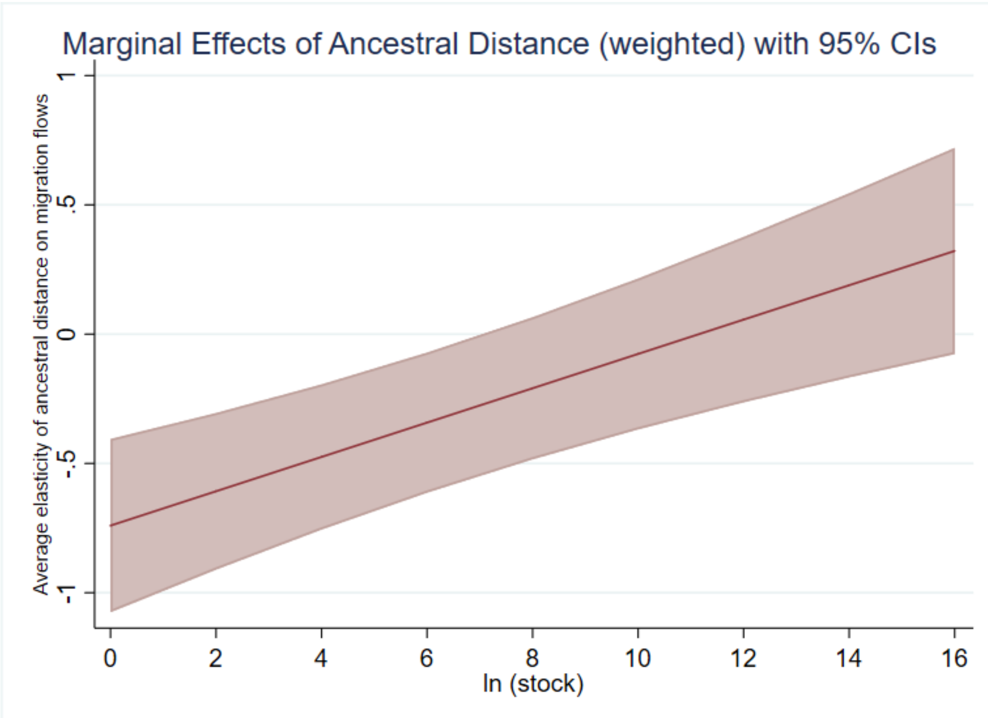
Reporting country	Criterion	Collection method
Argentina	COC	Residence permits
Australia	COB	Overseas arrivals and departures
Austria	COR	Population registers
Belgium	COC	Population registers
Brazil	COC	Residence permits and work authorisations
Canada	COB	Residence permits
Chile	COB	Residence permits
Czech Republic	COR	Population registers
Denmark	COR	Population registers
Finland	COR	Population registers
France	COC	Residence permits
Germany	COR	Population registers
Hungary	COC	Population registers and residence permits
Iceland	COR	Population registers
Israel	COB	Overseas arrivals and departures
Italy	COR	Population registers
Luxembourg	COC	Population registers
Mexico	COC	Residence permits and overseas arrivals
Netherlands	COB	Population registers
New Zealand	COC	Overseas arrivals and departures
Norway	COR	Population registers
Poland	COR	Population registers
Portugal	COC	Overseas arrivals and residence permits
Slovakia	COR	Population registers and residence permits
Slovenia	COC	Population registers
South Africa	COC	Residence permits
Spain	COR	Population registers
Sweden	COR	Population registers
Switzerland	COC	Residence permits
UK	COC	International Passenger Survey (IPS)
Uruguay	COC	Residence permits
USA	COB	Residence permits

Notes: COB: Country of Birth. COR: Country of last Residence. COC: Country of Citizenship.

2.7.3 Appendix A3. List of origin countries

Afghanistan, Albania, Algeria, Angola, Antigua & Barbuda, Argentina, Armenia, Australia, Austria, Azerbaijan, Bahamas, Bahrain, Bangladesh, Barbados, Belarus, Belgium, Belize, Benin, Bhutan, Bolivia, Botswana, Brazil, Brunei Darussalam, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, Cape Verde, Central African Republic, Chad, Chile, China, Colombia, Comoros, Republic of the Congo, the Democratic Republic of the Congo, Costa Rica, Cote d'Ivoire, Croatia, Cuba, Cyprus, Czech Republic, Denmark, Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Estonia, Ethiopia, Fiji, Finland, France, Gabon, Gambia, Georgia, Germany, Ghana, Greece, Grenada, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, Hungary, Iceland, India, Indonesia, Islamic Republic of Iran, Iraq, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kiribati, Democratic Peoples Republic of Korea, Republic of Korea, Kuwait, Kyrgyzstan, Lao Peoples Democratic Republic, Latvia, Lebanon, Lesotho, Liberia, Libyan Arab Jamahiriya, Lithuania, Luxembourg, Macedonia, the former Yugoslav Republic, Madagascar, Malawi, Malaysia, Mali, Malta, Mauritania, Mauritius, Mexico, Moldova, Mongolia, Morocco, Mozambique, Myanmar, Namibia, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Oman, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Qatar, Romania, Russian Federation, Rwanda, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Samoa, San Marino, Saudi Arabia, Senegal, Seychelles, Sierra Leone, Singapore, Slovakia, Slovenia, Solomon Islands, Somalia, South Africa, Spain, Sri Lanka, Sudan, Suriname, Swaziland, Sweden, Switzerland, Syrian Arab Republic, Tajikistan, Thailand, Tonga, Trinidad and Tobago, Tunisia, Turkey, Turkmenistan, Uganda, Ukraine, United Arab Emirates, United Kingdom, United States of America, Uruguay, Uzbekistan, Vanuatu, Venezuela, Vietnam, Zambia, Zimbabwe.

Figure 2.3: Marginal effects of Ancestral Distance on inflows at different levels of migration stocks from column (3) in Table 2.2.



2.7.4 Appendix A4. Marginal effects of Ancestral Distance on inflows

2.7.5 Appendix A5. Ancestral Distance (weigthed) matrix

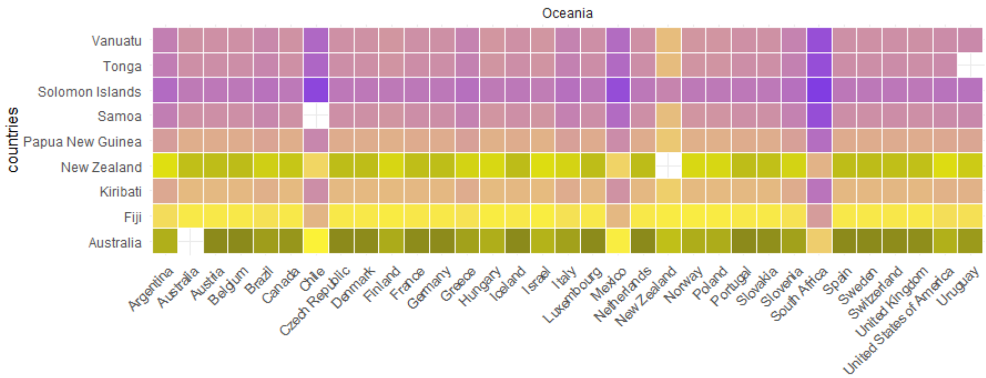


Figure 2.4: Ancestral Distance between countries - Oceania

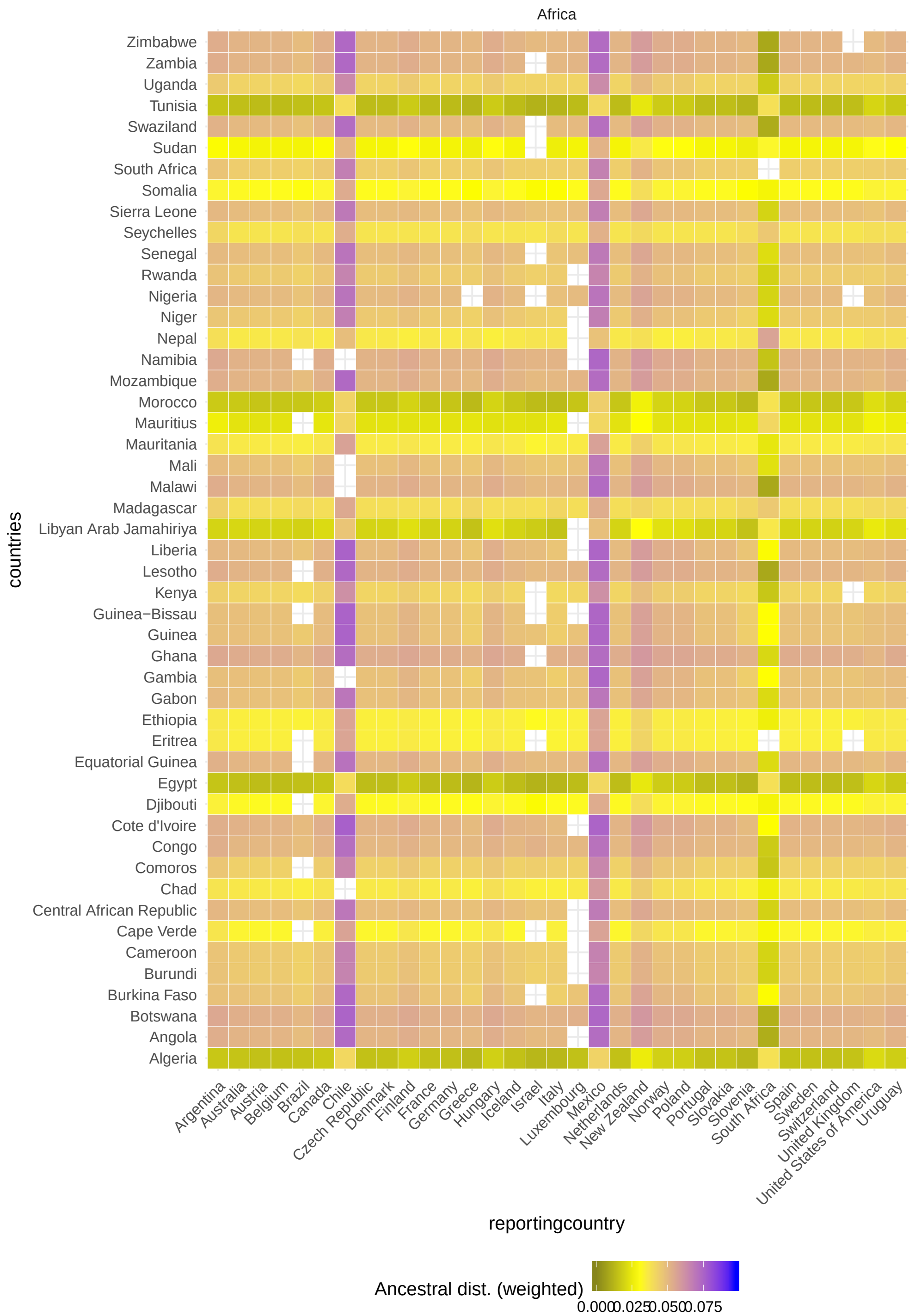


Figure 2.5: Genetic Distance between countries - Africa

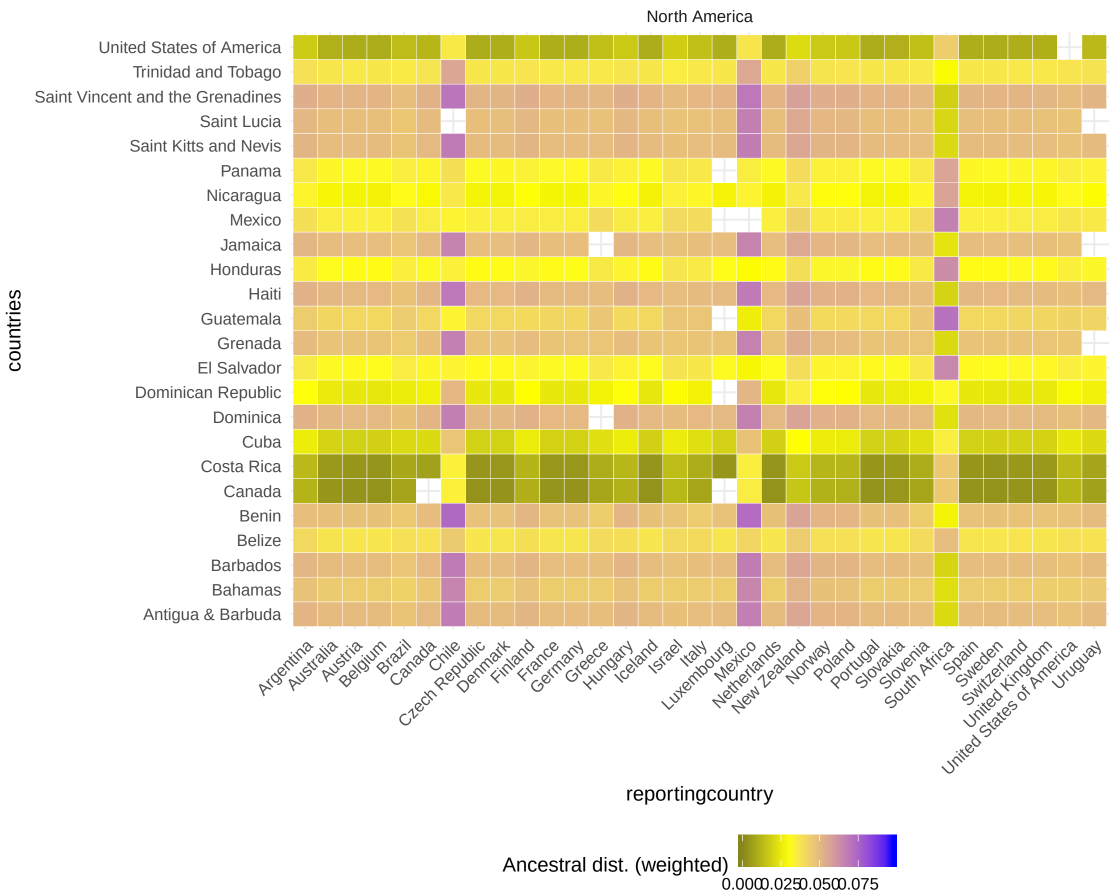


Figure 2.6: Genetic Distance between countries - North America

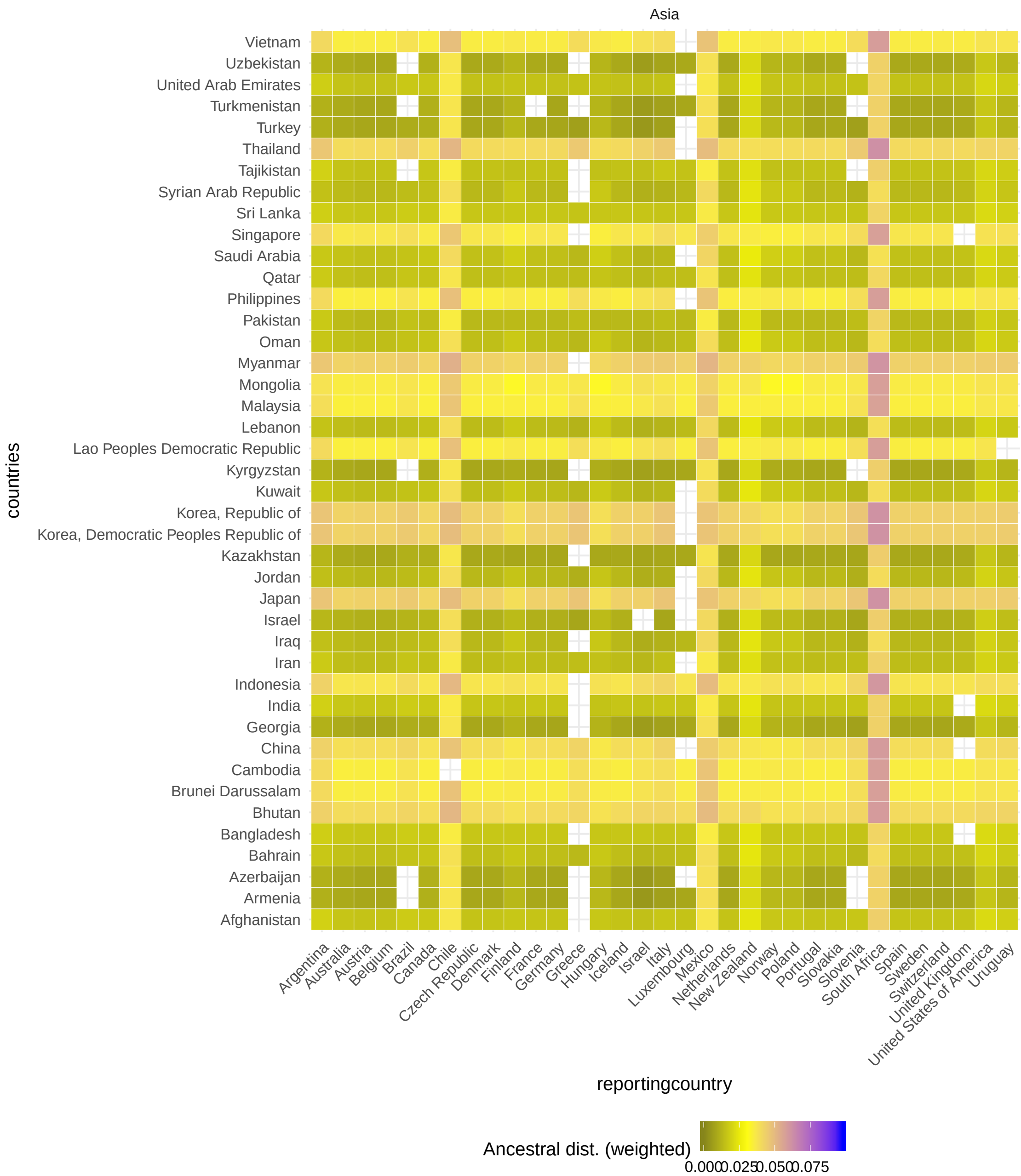


Figure 2.7: Genetic Distance between countries - Asia

Figure 2.8: Genetic Distance between countries - Europe

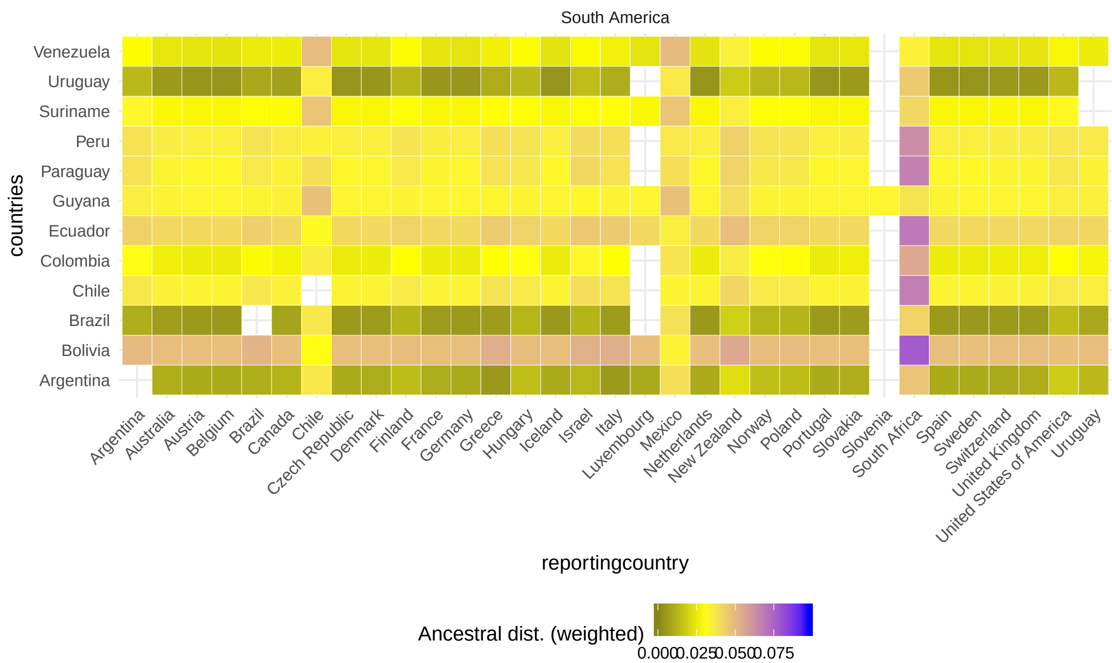


Figure 2.9: Genetic Distance between countries - South America

Appendix A6. PPML estimation of migration flows using ln (ancestral distance)

Table 2.9: PPML estimations of Migration Flows, ait and ajt FE

	(1)	(2)	(3)	(4)
log (stocks+1)	0.586***	0.580***	0.505***	0.630***
	(0.0209)	(0.0204)	(0.0353)	(0.0232)
log (AD*100+1)		-0.461**	-1.181***	
		(0.220)	(0.263)	
log (stocks+1) x log (AD*100+1)			0.0770***	0.0465**
			(0.0264)	(0.0213)
log (geodesic distance)	-0.308***	-0.296***	-0.284***	
	(0.0546)	(0.0576)	(0.0576)	
Contiguity	-0.270*	-0.277*	-0.238	
	(0.164)	(0.163)	(0.166)	
Colonial link	0.206*	0.241**	0.245**	
	(0.113)	(0.108)	(0.109)	
Common language	0.677***	0.669***	0.673***	
	(0.0904)	(0.0906)	(0.0910)	
Observations	17394	17394	17394	18229
$R*^2$	0.922	0.922	0.922	0.901
Dyadic FE ($\alpha_{i,j}$)	No	No	No	Yes
Origin-time FE ($\alpha_{i,t}$)	Yes	Yes	Yes	Yes
Destination-time FE ($\alpha_{j,t}$)	Yes	Yes	Yes	Yes

Notes: Ancestral dist. (weighted) has been multiplied by a 100
The standard errors are presented in parentheses clustered at the bilateral pair level. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CHAPTER 3
THE SOLUTION OF THE IMMIGRANT PARADOX: ASPIRATIONS AND EXPECTATIONS OF CHILDREN OF MIGRANTS

This paper is a joint work with Prof. Dr. Michel Beine and Prof. Dr. Skerdiladja Zanaj

3.1 Introduction

This chapter aims to determine the effect of educational aspiration and expectations on the school performance of children of migrants compared to children of natives parents in the United States. *Aspirations* describe personal goals such as the desired education level, the desired occupation, or career. Early studies, such as Bandura (2001) indicate that educational aspirations are key determinants of educational and professional life choices and outcomes. Aspirations are critical not only to explain individual performance but also the success of organizations. For instance, employees' career aspirations benefit enterprises by contributing to organizational capabilities and connections (Jung and Lee, 2019).

The process of setting educational goals depends on motivational components (e.g., desires) and also on contextual ingredients formed by family, school, and neighborhood characteristics (Sewell et al., 1969). Contextual components heavily influence the *expectations* of the educational goal. Unlike aspirations which are confined to desires, expectations embody the constraints that could influence aspects of the future. Both expectations and aspirations are as important as the ability to shape education outcomes (Lent et al., 1994).

Interestingly, aspiration and expectation may be aligned or misaligned. Alignment occurs when restrictions are perceived as possible to overcome. However, desires and expectations may be misaligned. For instance, a young individual can be highly inspirational by nature (e.g., dream to go to an Ivy League university), but she can confront various constraints (e.g., financial) that downgrade her expectations. The consequences of the misalignment between aspirations and expectations are not clear. On the one hand, very high aspirations with low expectations can lead to frustration and underachievement. On the other hand, reachable aspirations can inspire individuals (Genicot and Ray, 2017), regardless of their level of expectations. Experimental evidence from cognitive psychology and sports science show that goals that lie ahead but not too long ahead can be the best motivators (Berger and Pope, 2011; Latham and Locke, 1991). Which effect dominates on average is nevertheless not known and calls for some empirical analysis aiming at shedding light on the net impact on educational attainment.

Is there any systematic difference between migrant and native teenagers in the misalignment and its corresponding effects? In this chapter, we push forward the hypothesis that expectations and aspirations crucially affect the educational outcomes of young adults. Importantly, these effects are heterogeneous for natives and migrants. We hypothesize that the difference in performance between migrant children and natives lies within the misalignment between aspirations and expectations that migrant children form. We assess such a hypothesis using individual data of American pupils and show how the distance between the aspirations and expectations of the student affects their school performance.The students were asked in Wave I to rate on a scale of one to five, where one is low, and five is high, how much do they want to go to college, and how likely it is that they will go to college. Using these two questions, this chapter shows that the misalignment between aspirations and expectations is a driving force for better education outcomes of immigrant teenagers in the USA. This force can be so large that it fully explains the difference in educational outcomes between migrant and native children.

This chapter is directly related to an extensive large literature on the educational outcomes of immigrant children. The most accepted result is that in the USA, immigrant teenagers over-perform their corresponding native peers. This is known as the immigrant paradox (Palacios et al., 2008). Prior literature brings forward several mechanisms that could potentially explain this paradox, but none of these studies show which force completely explains such a paradox. We document the immigrant paradox, and we show that only considering family and school characteristics is not enough. Both aspirations and expectations matter, but they do so in a similar fashion for the natives as well as for the migrants. However, teenagers with an immigrant background express higher aspirations than expectations (they dream more than what they expect) when compared with their native peers. This misalignment is the key factor that motivates students to increase their efforts to study, leading to the paradox. Notwithstanding, when aspirations and expectations are aligned, the over-performance of migrant youngsters vanishes. Importantly, it is only the migrant children, i.e., pupils not born in the USA, that hold a strong misalignment and lead the immigrant paradox.

To uncover the above effects of aspirations, expectations, and misalignment, we use the restricted-use version of the National Longitudinal Study of Adolescent to Adult Health (Add Health) with a sample of 20,774 adolescents between the school grades 7-12 drawn from a representative sample of schools in the United States. Aspirations and expectations are constructed using two questions asked in Wave I. The first question pertains to their wish to go to college, measuring their educational aspirations, while the second one captures specifically to what extent they expect to go to college.[1] Using these questions, we measure misalignment between aspirations and expectations as the gap between the two measures. There exist two types of misalignment. Individuals may have high expectations but low desires, or on the contrary, they may exhibit low expectations but high aspirations. Our measure of educational performance is the overall final Grade-Point-Average (GPA). We also explore the final grade for Mathematics, English Literature, and Science. These grades are a strong predictor of the final test score (Scholastic Assessment Test-SAT) and posterior educational attainment.

Our results document the size of the effect of aspirations and expectations on educational outcomes of teenagers controlling for a very extensive list of contextual factors (family, cohort, school, neighborhood) and individual characteristics (personality traits, cognitive skill, BMI, among others). We show that greater aspirations are associated with a higher grade point average (GPA) for any subject at the end of high school. There is no differentiated effect of aspirations on the final GPA between children of migrants and natives. The results show that immigrant teenagers (the so-called 1.5 migrant generation) with low expectations but high aspirations are the ones driving the immigrant paradox. Clearly, the misalignment between what a teenager dreams and what she expects to be possible are key factors in explaining the performance at school. Our results show that the immigrant over-performance vanishes, and it is fully explained by the fact that these students have misaligned aspirations and expectations. The results are robust to the exclusion of students who migrated before the age of six or students who migrated after the age of 14.

Moreover, the results are robust to the exclusion of students with one migrant parent

[1] More precisely, students were asked: "rate on a scale of one to five, where one is low, and five is high, how much do you want to go to college? this question measures desires or aspirations. The second question is "how likely it is that you will go to college?"

and native parent that captures the potential advantages and disadvantages of inter-ethnic unions. Also, the results are consistent with the removal of teens who attended 12 grade during Wave 1 in order to ensure a strict temporal order between the dependent and independent variables. Furthermore, we test the potential importance of unobserved confounders by employing the formal approach proposed by Oster (2019). We show that potential omitted variable bias does not make our results statistically invalid.

The chapter is organized as follows. Section 2 provides selective coverage of the relevant literature to which our paper is connected. Section 3 details the data used to assess the impact of aspirations and expectations on school performance. Section 4 presents our econometric specification and discusses econometric issues related to endogeneity. Section 5 presents our benchmark results and auxiliary results, allowing us to uncover the mechanism at work to explain the impact of misalignment on school performance. It also presents a set of sensitivity analyses in terms of samples and identification of the effects. Section 6 briefly concludes.

3.2 Related Literature

This chapter contributes to the existing literature investigating the degree of integration of migrants by exploring the difference in academic performance between children in immigrant families and their non-migrant peers.[2] First-generation immigrants tend to exhibit lower educational attainment when compared to the native's counterparts (Dustmann and

[2]A different approach analyses the inter-generational transmission of human capital, comparing the educational attainment of the children compared to the education level of their parents. This strand of literature argues that the educational achievements of the children of migrants are strongly correlated with the educational achievement of their parent's generation (Dustmann and Glitz, 2011; Card et al., 2000; Gang and Zimmermann, 2000)), while the school system or the characteristics of the destination play a smaller role. This correlation, however, does not differ when compared to non-migrant populations (Smith, 2003).

Glitz, 2011). Yet, the successive generation of migrants in the USA is rapidly reverting this trend. In fact, children of migrants in the USA outperform children of natives from a similar socio-economic background in many educational indicators (Feliciano and Lanuza, 2017).[3] The immigrant paradox has been documented in previous academic work. To mention a few, Card (2005) and Chiswick and DebBurman (2004) show that the offspring of migrants born in the USA have achieved more years of education when compared to native individuals. These results are similar to those reported by Figlio et al. (2019) when analyzing test scores as a measure of academic performance. Figlio et al. (2019) found that, on average, children of migrants in Florida outperform white natives over time in both mathematics and reading. The performance advantage is predominantly observed in individuals with an Asian background (Portes and MacLeod, 1999; Feliciano, 2005) or from cultures that value long-term orientation (Figlio et al., 2019).

The immigrant paradox does not manifest only in the USA (Schnepf, 2007; Dustmann et al., 2012) but more broadly in English-speaking countries, whereas children of migrants performed better in reading and math test scores when it was measured in the Program for the International Assessment of Student Achievements - PISA (Schnepf, 2007).[4] Dustmann and Theodoropoulos (2010) finds that the educational attainment of British born minorities is higher than native British. Moreover, using the PISA database, Dustman et al. (2012) a negative gap in academic performance between children of migrants and natives living in countries such as Finland, Austria, Belgium, Netherlands, and Switzerland, even after controlling for family background, school characteristics and the share of migrants in the

[3]Moreover, a recent study by Abramitzky et al. (2019) show also an Immigrants' advantage in intergenerational mobility using data on millions of father-son pairs over a 100 years. According to their findings, children of migrants are more upwardly mobile than the children of US-born parents. This result shows that the children of migrants in the USA over-perform in terms of income mobility.

[4]However, in many other countries, migrant students lag behind native students (Riphahn, 2003; Algan et al., 2010)

school. In other OECD countries such as France, Greece, and Nordic countries (except for Finland), the gap in academic performance disappears after including a large set of control variables. Furthermore, Ours and Veenman (2003) compared second-generation migrants in the Netherlands with native Dutch people and showed that once age and parent's education are added as control variables, both groups do not show any significant difference in educational attainment.

An extensive list of explanatory factors has been proposed to understand the gap in academic performance between migrant students and natives. Factors such as selective migration policies (Levels et al., 2008; Entorf and Minoiu, 2005) parents' self-selection into migration (Feliciano, 2005), the social context at destination (Portes and Rumbaut, 1996; Portes and MacLeod, 1999), ethnic advantage (Borjas, 1992), and long-term orientation (Figlio et al., 2019) have been highlighted in the literature. Some authors have documented that ethnic minority adolescents express higher aspirations (Kao and Tienda, 1995) and higher expectations for university education (Boguslaw, 2017) when compared to native youth. To our knowledge, little attention has been paid to study teenager's attitudes such as aspirations and expectations as a potential explanation of the differences in school performance between migrants and natives. We fill precisely this gap. We contend that the individual attitudes and beliefs of youngsters are crucial ingredients that must enrich the knowledge of this phenomenon. Once controlling for a long list of individual, family, school, and neighborhood characteristics, we explore the effect of several and various individual characteristics and proxies of attitudes of USA's immigrant and native teenagers.

This chapter also contributes to the economic literature on the role of individual aspirations and goals on performance. Aspirations summarize preferences, a hope, or a wish to reach a goal, such as an occupation, obtaining a degree, or reaching a certain salary or wealth. Quaglia and Cobb (1996) defined aspirations as the "student's ability to identify and set goals for the future, while being inspired in the present to work toward those goals". In the economic literature, the concept of aspiration has been mostly addressed to study the poverty trap and its incidence in economic growth and inequality. For instance, in the theoretical papers of Dalton et al. (2016) and Genicot and Ray (2017), authors formalize the concept of aspirations as a reference point used by individuals. Deviations from the reference point are expressed as utility gains or losses from achieving an outcome (e.g., income). On the other hand, expectations, widely used in many fields of economics, reflect the constraints or beliefs acknowledged by an individual about aspects of the future. Expectations and the expected utility theory to analyze uncertain future events are ubiquitous in microeconomics and micro-founded macroeconomics from the seminal work of Morgenstern (1935) and Von Neumann and Morgenstern (1944). An individual uses his or her belief to create a probability distribution about the possible future scenarios. The terms aspirations and expectations are often used interchangeably and without precision. Notwithstanding, aspirations differ from expectations. The first concept represents ideals, while expectations embody constraints and perceived limitations (Böhme, 2015) or advantages. Therefore, aspirations and expectations can be aligned, but they can also be strongly misaligned.

How do aspirations affect future outcomes? Dalton et al. (2016) argue that there is two-way feedback between effort and aspirations. Individuals who do not internalize this relationship tend to aspire less than their actual capacity to achieve and to remain in the lower part of the income distribution. Under this theory, expectations about future outcomes are concealed under rational expectation equilibrium, where the expected value of the future income is equal to the future income. The consequences of the misalignment between aspirations and expectations are less clear. On the one hand, very high aspirations can lead to frustration and underachievement. While on the other hand, reachable aspirations can inspire individuals (Genicot and Ray, 2017). A recent study using data from India, Ross (2019) shows that the difference between the children's occupational aspiration and the current family status has an inverted U-shape relationship with human capital accumulation. Moreover, experimental evidence from cognitive psychology and sports science show that goals that lie ahead but not too long ahead can be the best motivators to improve performance (Berger and Pope, 2011; Latham and Locke, 1991). In this chapter, we first document the size of the effect of aspirations as well as expectations on education outcomes and then explore the effects of their misalignment. This exploration is missing in previous studies documenting the immigrant paradox.

3.3 Data and descriptive statistics

3.3.1 The Add Health data set

We employ the restricted-use version of the National Longitudinal Study of Adolescent to Adult Health (Add Health) collected by the Carolina Population Center. Add Health was designed to investigate the health, social conditions, education, environment, family situation, and friendships of adolescents in the United States of America throughout their transition into adulthood. While the study is not specifically devoted to migration questions, the sample size and the oversampling of particular migrant groups allow researchers to have a bigger sample size compared to other studies. The sample includes 20,774 adolescents between the grades 7-12 drawn from a representative sample of schools in the United States.[5] An

[5]To select the sample, all the students from each school filled a questionnaire at the school. The students were interviewed during the 1994-95 school year when they were between 13 and 18 years old. Using the in-school questionnaire, the Add Health researchers selected a random sample of students from strata defined by gender and grade (17 boys and 17 girls per grade per school).

extensive questionnaire was filled by the students at home. In addition, the parents of the students filled out a questionnaire that included questions about themselves, their partners, and the child.

The students were followed from 1994 until 2018 using five interviews. In Wave IV, subjects were aged between 24 and 32 years, when most of them had finished school and were entering the labor market. Our final sample of students consisted of 9,153 individuals. We omitted from the overall sample students with missing values (5,517 obs) in relevant questions. Wave III data contains follow-up interviews from the 14,979 initial respondents, which implies we do not observe the School transcript data for over 6,000 respondents from Wave I. We use adjusted sampling weights calculated by the Add Health team to account for panel attrition as well as school transcript non-response.[6]

3.3.2 Main variables

3.3.2 Outcome variables: educational outcomes

We studied school performance measured by the weighted average Grade-Point-Average (GPA) for Mathematics, English Literature, Science, and the overall GPA during the four years of high school. Although previous studies have used standardized test scores, we use grades since they are a strong predictor of the final test score (Scholastic Assessment Test-SAT) and posterior educational attainment. For example, Zwick and Sklar (2005) show that an increase in one standard deviation in high school GPA increases the first-year grade-point averages (FGPAs) among first-year college students by one-third standard deviation.

The GPA measures came from the Adolescent Health and Academic Achievement (AHAA)

[6]The transcripts were not collected when the respondent was home-schooled, attended high school outside the USA, the school closed, refused to provide information or the information was incomplete or incorrect.

study. The AHAA data corresponds to a collection of the school transcripts for 12,241 Add Health respondents from Wave I. The data allowed us to measure the performance of the students at the end of high school.

3.3.2 Aspirations, expectations and misalignment

We define aspirations as hopes and desires about the future, while expectations are the beliefs about what will happen in the future (DeMoss, 2013; Jacob and Wilder, 2010). To measure educational expectations and aspirations to attend college, we used two questions asked in Wave I, well before the measurement of the school performance. More specifically, questions about aspirations and expectations were asked in Wave I (1994-95), whereas the records of the school performance are taken from the school transcripts at the end of high school. Students were asked: "rate on a scale of one to five, where one is low, and five is high, how much do you want to go to college? And how likely it is that you will go to college"? We define three categories for each variable. Students with Low aspirations-expectations corresponds to those who answered on a scale of one to three, Medium corresponds to answering 4, and High corresponds to the maximum level.

We define misalignment as the difference between the level of aspiration and expectations shown by the students. We define three dummy variables: Asp $<$ Exp, equal to 1 is aspirations are smaller than expectations and zero otherwise; Asp $>$ Exp equal to 1 if aspirations are larger than expectations and zero otherwise; and Asp $=$ Exp equal to 1 if aspirations are equal to the level expectations and zero otherwise. When Asp$<$ Exp is equal to one, the adolescent is calculative, and she expects to go to college more than she desires it. By contrast, when Asp$>$ Exp is equal to one, the adolescent is inspirational, but she expects it to be difficult to attend college; thus, she downgrades her expectations. This second type of misalignment is prone to feelings of frustration that harm educational performance or, on the contrary, can be the driving force to better performance. We investigate which effect dominates in the determination of school performance and whether there is a heterogeneous effect for native and immigrant children.

3.3.2 Migration generation

We define 1.5 generation as those children who were not born in the USA and whose biological parents were born outside the USA. We use this definition following the work of Rumbaut (2004), in which the author identifies as the 1.5 generation as those migrants who arrived at the destination country between the ages of 6 and 12 years old and who have learned the origin's language but has completed an important part of their education at the destination country. Since the students in the Add Health sample were still in high school, many of them spent most of their school years in the US.[7] For this group of immigrants, migration is not a choice. Importantly, they lived the trauma of the migratory process with their parents, bringing with them some of the experiences accumulated in their countries of origin.

Table 3.1: Definition of migration generation

	Child	
Parent	Born in U.S	Born outside U.S
Born in U.S.	Natives	Natives
Born outside U.S.	Generation 2.0	Generation 1.5

The second-generation or generation 2.0 corresponds to children who were born in the USA but for which at least one of the biological parents was born outside the USA. Lastly,

[7]The average age of migration of the teens born abroad is 7.6 years old.

we considered native children such as those who were born in the USA and both of their biological parents as well. Children born abroad whose parents were born in the USA are also assimilated to native children.[8] Table 3.1 summarizes the different cases.

To determinate the migration generation, we used the country of birth indicated by the child. Nevertheless, in cases where the information was missing, we used the parent's response or the answers in the school questionnaire. In addition, we used the country of birth of the biological parents answered by the child in the questionnaire collected at home during Wave I. We also employed the answers from the questionnaire collected at the school when information about the biological parents was missing. When biological parents were absent, we used the information on the adoptive or step-parents.

3.3.2 Covariates

A comprehensive list of the control variables is available in Tables 3.15 and 3.16 reported in the appendix of this chapter. Among others, we control for cognitive ability and different non-cognitive traits that affect human capital investment, such as internal locus of control and self-esteem. According to Coleman and DeLeire (2003), teenagers who believe that outcomes are a result of their efforts have a larger likelihood of graduating from high school.Moreover, high self-esteem and confidence are associated with better learning and school success (Mocan and Yu, 2017). We included a self-esteem index constructed using different questions asked to the student in Wave I (see Table 3.15 for detail). Moreover, body mass index (BMI) is included in the regressions as a health indicator, but it also captures aspects of self-esteem development (Mocan and Tekin, 2011; Zuppann and Liu, 2016). In addition, we included age since it allowed us to take into account whether the students began high school at different

[8] Among the Add Health total sample, we have identified only 140 students born abroad and whose parents are US-born. Out of these 140 students, 40% of them migrated back to the USA before the first year.

ages. More mature students might have a better understanding of their aspirations and how to accomplish them. We also include gender and ethnicity as controls.

In line with the broad literature on education, we include other household controls such as family structure, number of siblings, parental expectations for higher education, parental involvement, income,[9] and a dichotomous variable that is equal to one if the family speaks English at home and zero otherwise. We include the education level of the highest educated parent. When the father is not present in the household, we use the education of the mother or the adult in charge.[10]

3.3.3 Descriptive Statistics

Tables 3.2 and 3.3 report the means and standard deviations for all control variables by type of pupils. Children of immigrants represent 21 percent of the sample. Migrant generation 2.0 and 1.5 represent 13 and 7 percent, respectively. Both children of immigrants and natives in this sample express strong desires to achieve a college education. While 73 percent of native children report the highest level of aspiration, nearly 79 percent belonging to the 1.5 generation express the same wish. High aspirations and middle expectations are predominant among the 1.5 generation children. Only 56 percent of them report having the same level of aspirations and expectations, while 38 percent report having larger aspirations than expectations. In contrast, approximately 22 percent of native children reported larger aspirations than expectations. Immigrant children might understand the benefits of higher education; however, they might perceive lower returns as a result of potential labor market

[9]There are missing values in family income because some parents were not surveyed in Wave I. Only 76% of the families reported income in the survey; therefore, we imputation some values using the mean of the income.

[10]27 percent of children do not report a father living in the household nor their education level. For a detailed description of the control variables see Table 3.15 and 3.16.

discrimination or the lack of role models in their community or neighborhood. This group of students reports a lower score in the vocabulary test (PPVT), lower self-confidence score, and lower body mass index when compared to the native students. Moreover, they were raised in families with lower incomes, more siblings, they are less likely to speak English at home, and the mothers have a lower education level when compared to natives. The proportion of 1.5 generation immigrant students whose mother did not finish high school is more than 36 percent, as opposed to only 10 percent of the native students.

The children corresponding to the second-generation immigrants show similar aspiration levels when compared to natives. However, over 27 percent of them show larger aspirations than expectations. Second-generation migrants do not seem to differ when compared to natives in aspects such as BMI, age, gender, internal locus of control, family income, or the number of siblings in the household. In contrast, they show significantly lower average scores in the vocabulary test. This could be explained by the fact that 29% live in families that do not speak English at home, and the mothers are less educated than the mothers of native students.

Despite these socioeconomic differences, the parents of both 1.5 and 2.0 generations of immigrants express high expectations for their child's academic future when compared to native adolescents. While 40 percent of native students have parents who express high expectations for college attendance for their children, this proportion is equal to 71 and 58 percent for the 1.5 and 2.0 generation of immigrants, respectively.

3.4 Empirical Strategy

We bring the data to econometric specifications using variation across individuals i, schools s, and education grades g. We estimate three specifications. The first capturing the role of

Table 3.2: Descriptive statistics for the Add Health sample - Part 1

	Natives	Gener. 1.5	Gener. 2.0	Mean Difference	
	(1)	(2)	(3)	(1)-(2)	(1)-(3)
College Aspirations					
(1-3)	0.142	0.091	0.105	0.050^{+}	0.036^{+}
	(0.336)	(0.345)	(0.372)	[0.019]	[0.016]
4	0.126	0.116	0.130	0.010	-0.004
	(0.320)	(0.383)	(0.408)	[0.021]	[0.014]
5	0.731	0.792	0.763	-0.060^{+}	-0.031
	(0.427)	(0.485)	(0.515)	[0.028]	[0.021]
College Expectations					
(1-3)	0.200	0.196	0.171	0.004	0.028
	(0.385)	(0.475)	(0.457)	[0.030]	[0.019]
4	0.202	0.315	0.249	-0.113^{+}	-0.046**
	(0.387)	(0.556)	(0.524)	[0.024]	[0.022]
5	0.596	0.487	0.579	0.017^{+}	0.017
	(0.472)	(0.598)	(0.598)	[0.038]	[0.273]
Alignment					
Asp.= Exp.	0.711	0.560	0.650	0.150^{+}	0.061**
	(0.436)	(0.593)	(0.578)	[0.034]	[0.022]
Asp.< Exp.	0.068	0.0503	0.072	0.017	-0.004
	(0.242)	(0.26)	(0.313)	[0.012]	[0.011]
Asp.> Exp.	0.220	0.389	0.277	-0.168**	-0.056**
	(0.399)	(0.583)	(0.542)	[0.035]	[0.021]
Age	15.38	15.88	15.41	-0.502**	-0.028
	(1.705)	(2.085)	(2.19)	[0.238]	[0.168]
Male	0.499	0.454	0.493	-0.044	-0.005
	(0.481)	(0.595)	(0.605)	[0.031]	[0.026]
Body-Mass-Index	22.467	21.709	22.301	0.758^{+}	0.166
	(4.379)	(4.551)	(5.259)	[0.271]	[0.223]
PPVT	103.862	89.986	102.558	13.875^{+}	1.304
	(12.606)	(20.144)	(17.139)	[1.476]	[1.006]
Self-esteem index	0.186	-0.322	0.018	0.509**	0.1683
	(1.808)	(2.193)	(2.592)	[0.119]	[0.118]
Internal Locus =1	0.753	0.781	0.739	-0.028	0.013
	(0.415)	(0.494)	(0.532)	[0.030]	[0.021]
Observations	7356	643	1154		

Notes: Standard deviations are in parentheses and standard errors are in brackets. PPVT: Peabody Picture Vocabulary Test. * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$.

Table 3.3: Descriptive statistics for the Add Health sample - Part 2

	Natives	Gener. 1.5	Gener. 2.0	Mean difference	
	(1)	(2)	(3)	(1)-(2)	(1)-(3)
English at Home =1	0.996	0.316	0.705	0.680^{+}	0.291^{+}
	(0.059)	(0.556)	(0.552)	[0.055]	[0.036]
White non-hispanic	0.770	0.088	0.324	0.681^{+}	0.445^{+}
	(0.405)	(0.339)	(0.567)	[0.032]	[0.037]
Hispanic	0.045	0.482	0.413	-0.437^{+}	-0.367^{+}
	(0.200)	(0.598)	(0.596)	[0.077]	[0.043]
Black non-Hispanic	0.170	0.039	0.063	0.131^{+}	0.107^{+}
	(0.362)	(0.232)	(0.295)	[0.025]	[0.020]
Asian	0.007	0.373	0.152	-0.365^{+}	-0.145^{+}
	(0.081)	(0.578)	(0.435)	[0.068]	[0.023]
Other	0.007	0.020	0.045	-0.010	-0.039**
	(0.080)	(0.573)	(0.252)	[0.010]	[0.014]
N siblings	1.435	2.406	1.825	-0.970^{+}	-0.389
	(1.189)	(2.516)	(1.711)	[0.208]	[0.119]
Mother Education					
Less than high school	0.103	0.368	0.234	-0.264^{+}	-0.130^{+}
	(0.296)	(0.574)	(0.502)	[0.048]	[0.035]
High school graduate	0.608	0.259	0.462	0.349^{+}	0.145^{+}
	(0.473)	(0.521)	(0.591)	[0.033]	[0.029]
College graduate	0.287	0.2535	0.236	0.033	0.051*
	(0.439)	(0.517)	(0.503)	[0.048]	[0.029]
Missing information		0.118	0.066	-0.118+	-0.066+
		(0.384)	(0.295)	[0.031]	[0.014]
Both biological parents	0.614	0.652	0.702	-0.037	-0.087**
	(0.472)	(0.566)	(0.542)	[0.047]	[0.028]
At least one step-parent	0.159	0.1364	0.1384	0.023	0.0210
	(0.355)	(0.408)	(0.409)	[0.022]	[0.020]
Single parent or other	0.225	0.211	0.158	0.014	0.066
	(0.405)	(0.485)	(0.433)	[0.040]	[0.021]
High Parents expectations	0.406	0.715	0.584	-0.292^{+}	-0.182^{+}
	(0.476)	(0.536)	(0.584)	[0.028]	[0.029]
Parent involvement index	0.029	0.163	-0.014	0.013	0.043
	(1.267)	(1.421)	(1.402)	[0.093]	[0.065]
Contextual educational	36.9653	79.935	72.575	-42.970	-35.609
attainment	(15.110)	(13.692)	(22.0149)	[1.091]	[0.944]
Income (Thousand)*	48.838	32.524	46.487	16.314^{+}	2.351
	(43.803)	(45.816)	(55.858)	[4.153]	[3.140]
Observations	7356	643	1154		

Notes: Standard deviations are in parentheses and standard errors are in brackets.
* Income is reported for 7,103 respondents * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$.

aspirations and expectations and their interaction with migration generation. The second estimates the effect of misalignment and its interaction with migration generation, and the third includes misalignment and aspirations and expectations as additional control variables:

$$
\begin{aligned}
Y_{isg,t+4} &= \alpha_0 + \alpha_1 Gen_{isg} + \alpha_2 High_{isg,t} + \alpha_3 Med_{isg,t} + \alpha_4 Gen_{isg} \times High_{isg,t} + \\
&\quad + \alpha_5 Gen_{isg} \times Med_{isg,t} + \alpha_6 X_{isg,t} + \mu_s + \mu_g + \epsilon_{isg,t+4}
\end{aligned} \tag{3.1}
$$

$$
\begin{aligned}
Y_{isg,t+4} &= \beta_0 + \beta_1 Gen_{isg,t} + \beta_2 W_{isg,t} + \beta_3 D_{isg,t} + \beta_4 Gen_{isg} \times W_{isg,t} \\
&\quad + \beta_5 Gen_{isg} \times D_{isg,t} + \beta_6 X_{isg,t} + \mu_s + \mu_g + \epsilon_{isg,t+4}
\end{aligned} \tag{3.2}
$$

$$
\begin{aligned}
Y_{isg,t+4} &= \gamma_0 + \gamma_1 Gen_{isg} + \gamma_2 High_{isg,t} + \gamma_3 Med_{isg,t} + \gamma_4 W_{isg,t} + \\
&\quad \gamma_5 D_{isg,t} + \gamma_6 Gen_{isg} \times W_{isg,t} + \gamma_8 Gen_{isg} \times D_{isg,t} + \gamma_9 X_{isg,t} + \mu_s + \mu_g + \epsilon_{isg,t+4}
\end{aligned} \tag{3.3}
$$

where $Y_{isg,t+4}$ is either the total average GPA and the GPA in Math, Science, or English literature at the end of high school. Gen_{isg} is a vector of binary variables indicating whether the teenager is a 1.5 generation migrant, a second-generation migrant, or a native. The dummy variables $High_{isg,t}$ correspond to the cases where aspirations or/and expectations are high. $Med_{isg,t}$ correspond to the cases where aspirations or/and expectations have a medium level. The variable $W_{isg,t}$ is equal to 1 for the cases where Asp$<$ Exp and zero otherwise. The variable $D_{isg,t}$ is equal to 1 for the cases where Asp$>$ Exp, and zero otherwise. The baseline category is when there is no difference between the level of aspiration and expectations chosen by the students. We include school and grade fixed effects (μ_s and μ_g respectively). We also substitute the school fixed effects with neighborhood fixed effects as an alternative since not all schools are nested within the same neighborhood or vice versa. [11] Moreover, the sample size within each school limits the use of a cross-classified multilevel model.

3.4.1 Endogeneity issues

Endogeneity constitutes a threat to the identification of the causal effect of aspirations and expectations on academic attainment. First, individuals may form their aspirations based on their performance and vice-versa. Similarly, children can update their expectations as a response to their performance. This would create a case for reverse causality, inducing a bias in the estimates of the effect of these variables on education attainment. Nevertheless, the design of the survey mitigates this concern as the expectations and aspirations of the children are measured well before the measurement of their academic performances. In this sense, aspirations, expectations, and possible misalignment are predetermined with respect to performance.[12]

The second source of endogeneity is related to the omission of unobserved factors that determine both aspirations and expectations on the one hand, and academic performance on the other hand. While the inclusion of school and grade fixed effects mitigates the role of these unobserved common determinants, the fact that not all relevant questions were asked throughout the different waves of the survey prevents us from including individual fixed effects.[13] In the literature, additional solutions to deal with the remaining concern of en-

[11] The Add Health sample includes charter, choice, and magnet schools that offer open enrollment programs allowing students to attend schools outside their residence districts.

[12] We also observe that expectations and aspirations do not vary for most students when we compare the responses to the survey in Wave I and II.As shown in figure 3.1 in Appendix B, over 60% of the respondents do not change their level of aspirations between Wave I and Wave II. For the case of expectations, over 50% of the respondents do not modify their answer between Wave I and Wave II.

[13] We control for a large set of individual-specific variables such as age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, and household income.

dogeneity have been used, ranging from experimental designs to "peer-effects" instrumental variables.[14] In our context, given the richness of the included covariates, we did not find "peer-effects" type of instruments that comply with the conditions of validity of instrumental variables.

As an alternative to experimental approaches or instrumental variable solutions, we evaluate the robustness of the results by analyzing the stability of the coefficient of interest to the inclusion of observed controls employing the formal approach proposed by Oster (2019). The idea is to evaluate how important is the size of the omitted variables to invalidate the obtained estimates, making assumptions about the relationship between selection along with observable and unobservable determinants. Based on Altonji et al. (2005), Oster (2019) presents the connection between omitted bias and coefficient stability theoretically by exploiting the coefficient stability and R-squared movements. Formally, Oster (2019) proposes the following adjusted coefficient of interest:

$$\gamma_{adjusted} = \tilde{\gamma} - \delta[\gamma^{*} - \tilde{\gamma}]\frac{R_{max} - \tilde{R}}{\tilde{R} - R^{*}} \tag{3.4}$$

where $\tilde{\gamma}$ and $\tilde{R}$ correspond to the coefficient of interest and the R^2 from the regression with controls. γ^* and R^* correspond to the coefficient and the R^2 from the regression without controls. R_{max} would be the maximum possible R^2 if both unobserved and observed

[14]Prior empirical literature has relied predominantly on randomized control trials (Goux et al., 2017; Carlana et al., 2018; Bernard et al., 2014, 2019) Researchers have conducted different experimental interventions to influence the individual's level of aspirations such as videos, meetings with school principals, or counseling programs for either parents or teenagers. Christofides et al. (2015), alternatively, use an instrumental variable approach to determine the impact of aspirations on educational outcomes such as going to university and graduating. The authors use as an instrumental variable the change in the student's belief about whether a university degree is required to work in the future job at age 30. Moreover, recent studies have made use of "peer-effects" instrumental variables. For example, Kosec et al. (2018) employ as an instrumental variable a predicted aspiration index using five dimensions of aspirations and the average weights of the community instead of the individual's weight in the construction of the index. To our knowledge, no studies have used these two methodologies to estimate the effect of the gap between educational aspirations and expectations.

variables were included in the specification. A maximum value of R_{max} would be 1, while a minimum value would be $\tilde{R}$. The parameter δ corresponds to the degree of selection on unobserved factors proportional to the observable characteristics necessary to make the coefficient of interest statistically insignificant ($\gamma = 0$). Oster (2019) proposes two approaches for robustness. The first, in which the researcher assumes a value for R_{max} and calculates the relative degree of selection on unobservables proportional to observable factors (δ) for which $\gamma = 0$. The second, in which the researchers use bounds on R_{max} and δ to develop a set of bounds for γ. While this method relies on the assumption that the relationship between non-observable factors and the treatment can be retrieved from the relationship between the observable variables and the treatment, it is informative about the degree of omitted variable bias in our results. We adapt this framework to ours in which the treatment effect varies with aspirations/expectations. As shown in Section 3.5.3, reassuringly, the problem of omitted variable bias seems negligible in our estimations.

3.5 Results

The regression results for the overall GPA are presented in Table 3.4. Tables 3.5, 3.6 and 3.7 report the results when the dependent variable is the GPA in mathematics, science, or English literature, respectively. These tables include all the control variables reported in Tables 3.2 and 3.3 as well as grade and school fixed effects.[15] Since we study three variables, aspiration, expectations, and the misalignment between aspirations and expectations, the tables report three different specifications. Column (1) reports the results where we regress GPA on migration generation, including all control variables. Columns (2), (3), and (5) display the

[15]For space considerations, the estimated coefficients for the control variables are not reported but are available upon request.

interaction between migration generation and the student's aspirations, expectations, and misalignment, respectively. Finally, column (6) includes the interaction between migration generation and the student's misalignment while aspirations and expectations are added as control variables.

According to the baseline estimates, children born abroad have a higher overall GPA score than native children after controlling for a very extensive list of individual, family, and school fixed effects. The difference in the GPA between 1.5 immigrant teenagers and native teens is only 0.16 points, as depicted in row one in column (1) of Table 3.4. Moreover, after including an extensive list of control variables, we found no significant differences in GPA between US-born teens of a migrant parent (2.0 migrants) and the US-born teens of US-born parents.

Different studies illustrate that immigrant parents and their children express high educational aspirations and expectations (Kao and Tienda, 1995; Tjaden and Hunkler, 2017; Tjaden and Scharenberg, 2017). While the descriptive statistics' session shows that 1.5 generation migrant teens do express higher aspirations to attend college when compared to native teens, we found no differentiated effect of aspirations on the final GPA between the children of migrants and the children of natives (see Table 3.4 column 3) when we control for different covariates. A similar pattern was found when analyzing specific subjects as shown in column 3 from Tables 3.5, 3.6 and 3.7. Results show that greater aspirations are associated with a higher grade point average (GPA) for any subject at the end of high school. This pattern is general for all the interviewed teens.

Meanwhile, for the case of expectations, we also found a positive association between higher expectations to attend college and high school final GPA. Nevertheless, we also found a negative and significant interaction effect between high-level expectation and being for

Table 3.4: OLS regression results for Overall GPA (4 year average) with school fixed effects

	(1) gpal	(2) gpal	(3) gpal	(4) gpal	(5) gpal	(6) gpal
Gener. 1.5	0.168**	0.314*	0.476+	0.188+	0.0981	0.0789
	(0.0662)	(0.185)	(0.117)	(0.0672)	(0.0751)	(0.0756)
Gener. 2	-0.0408	-0.0161	0.0181	-0.0279	0.000155	0.00304
	(0.0523)	(0.135)	(0.104)	(0.0525)	(0.0571)	(0.0543)
Medium aspi.		0.151+				0.128**
		(0.0425)				(0.0560)
High aspi.		0.349+				0.288+
		(0.0346)				(0.0852)
Gener. 1.5 × Medium aspi.		-0.0630				
		(0.228)				
Gener. 1.5 × High aspi.		-0.205				
		(0.188)				
Gener. 2 × Medium aspi.		-0.0744				
		(0.136)				
Gener. 2 × High aspi.		-0.0193				
		(0.130)				
Medium Exp.			0.208+			0.0933*
			(0.0343)			(0.0482)
High Exp.			0.439+			0.200**
			(0.0335)			(0.0887)
Gener. 1.5 × Medium Exp.			-0.224*			
			(0.127)			
Gener. 1.5 × High Exp.			-0.438+			
			(0.119)			
Gener. 2 × Medium Exp.			-0.0609			
			(0.113)			
Gener. 2 × High Exp.			-0.0329			
			(0.108)			
Asp.< Exp.				-0.0719*	-0.0568	0.0800
				(0.0382)	(0.0419)	(0.0630)
Asp.> Exp.				-0.148+	-0.160+	-0.119*
				(0.0286)	(0.0320)	(0.0645)
Gener. 1.5 × Asp.< Exp.					-0.0630	-0.0160
					(0.143)	(0.131)
Gener. 1.5 × Asp.> Exp.					0.254+	0.264+
					(0.0951)	(0.0957)
Gener. 2 × Asp.< Exp.					-0.151	-0.151
					(0.103)	(0.102)
Gener. 2 × Asp.> Exp.					-0.0494	-0.0437
					(0.0892)	(0.0859)
Constant	2.984+	2.883+	2.826+	2.974+	3.013+	2.839+
	(0.421)	(0.421)	(0.405)	(0.417)	(0.412)	(0.401)
Observations	9153	9153	9153	9153	9153	9153
R^2	0.397	0.415	0.427	0.401	0.403	0.430
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$. The results in this table were estimated using Wave III Education Sample weights.

1.5 migrant generation children when compared to natives. This means that while higher expectations at the start of high school are important, it seems that at higher levels of expectation, the difference in GPA between native teenagers and migrant teenagers decreases. When both groups of students have high expectations, the difference in the average GPA is closer to zero (0.038). When the students have a medium expectation level, the difference in the average overall GPA between native and 1.5 children is 0.25 points. When the students have a low expectation level, the difference in the average overall GPA between native and 1.5 children is 0.476. A similar pattern is found when the dependent variable is a specific subject such as Math, English literature or Science (see column 3 in Tables 3.5, 3.6 and 3.7). These results lead us to explore the gap between aspirations and expectations.

As explained by Genicot and Ray (2017), the absolute level of aspirations is not enough to explain performance. Instead, researchers must consider the distance between the aspiration and the status quo (or the perception of the status quo in this case) to understand how individuals manage to achieve the aspired goal. Following this theoretical result, we estimate the effect of misalignment between aspirations and expectations on the difference in GPA between 1.5 migrant children and natives. We uncovered that misalignment between aspiration and expectations is associated with lower grades for most children. When all groups of students have aligned aspirations and expectations, the difference in the GPA between 1,5 migrant generation, second generation, and native children are statistically insignificant. However, there is a positive and significant interaction effect between frustration (i.e., high aspiration but low expectation) and being a 1.5 migrant teen. When the teens express higher aspirations than expectation, the difference in the average overall GPA between native and 1.5 children is 0.26 points, which is equivalent to a difference of 0.31 standard deviations. While this difference is small, it also suggests that migrant children might have a positive reaction when facing misaligned aspirations that reflects in their final high school grades.

We find similar results using neighborhood fixed effects (See tables 3.21 to 3.24 in Appendix B). Interestingly, it appears that 1.5 migrants who are endowed with educational aspirations but are pessimistic about their future educational career do not give up their dreams. The results point to conclude that teenagers with low expectations, but possibly high aspirations, are the ones explaining the positive difference in GPA between migrant children and natives. It is precisely this difference the driving force of the immigrant paradox. As we will explore in Section 3.5.1, this frustration makes 1.5 generation migrant children spend less time in leisure activities and possibly study more.

3.5.1 Mechanism: role of misalignment on effort and leisure

In this section, we study a potential mechanism that links student goals and beliefs with outcomes. We explore the idea that migrant children might compensate for their perceived disadvantage with an increase in their studying effort. Misalignment can be a driving force to study more rather than disappointment and giving up. To test this hypothesis, we estimate auxiliary regressions and introduce an outcome variable that measures the number of hours teenagers spend watching TV as a proxy of leisure time and a possible direct substitute for studying time.[16] The Column (2) in Table 3.8 reports estimates of the relationship between migrant generation and (mis)alignment on the number of hours watching TV measured in Wave II. Column (3) includes the level of hours watching TV measured in Wave I. In general, we found that migrant children who have misaligned aspirations in Wave 1 are associated

[16]Other potential activities could have also been considered as leisure activities such as playing video games. However, not all children likely possessed a video console in 1994, and it might also reflect some income differences. The survey does not give information about the hours spent doing homework or playing sports. Nevertheless, it is debatable whether sports should be a direct substitute to study time since sport can improve the health of children and make them more able to performs other tasks, including school activities.

Table 3.5: OLS regression results for Math GPA (4 year average) with school fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
Gener. 1.5	0.172*	0.291	0.470+	0.191**	0.0779	0.0625
	(0.0900)	(0.216)	(0.145)	(0.0906)	(0.0958)	(0.0971)
Gener. 2	-0.0157	0.0714	-0.00673	-0.00485	0.0323	0.0351
	(0.0620)	(0.136)	(0.107)	(0.0618)	(0.0663)	(0.0642)
Medium aspi.		0.112**				0.0888
		(0.0464)				(0.0665)
High aspi.		0.289+				0.245**
		(0.0422)				(0.107)
Gener. 1.5 × Medium aspi.		0.0332				
		(0.262)				
Gener. 1.5 × High aspi.		-0.181				
		(0.210)				
Gener. 2 × Medium aspi.		-0.188				
		(0.149)				
Gener. 2 × High aspi.		-0.0806				
		(0.139)				
Medium Exp.			0.148+			0.0594
			(0.0439)			(0.0599)
High Exp.			0.381+			0.169
			(0.0397)			(0.112)
Gener. 1.5 × Medium Exp.			-0.102			
			(0.156)			
Gener. 1.5 × High Exp.			-0.491+			
			(0.139)			
Gener. 2 × Medium Exp.			0.0398			
			(0.131)			
Gener. 2 × High Exp.			0.0103			
			(0.115)			
Asp.< Exp.				-0.0219	-0.00350	0.126
				(0.0508)	(0.0565)	(0.0824)
Asp.> Exp.				-0.131+	-0.146+	-0.0994
				(0.0380)	(0.0423)	(0.0876)
Gener. 1.5 × Asp.< Exp.					-0.0954	-0.0567
					(0.142)	(0.137)
Gener. 1.5 × Asp.> Exp.					0.320+	0.328+
					(0.116)	(0.118)
Gener. 2 × Asp.< Exp.					-0.178	-0.175
					(0.116)	(0.117)
Gener. 2 × Asp.> Exp.					-0.0734	-0.0676
					(0.106)	(0.104)
Constant	2.869+	2.795+	2.752+	2.853+	2.902+	2.766+
	(0.485)	(0.489)	(0.471)	(0.480)	(0.473)	(0.466)
Observations	9124	9124	9124	9124	9124	9124
R^2	0.280	0.290	0.299	0.283	0.284	0.300
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. Significance levels * $p < 0.10$, ** $p < 0.05$, + $p < 0.01$. The results in this table were estimated using Wave III Education Sample weights.

Table 3.6: OLS regression results for English GPA (4 year average) with school fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
Gener. 1.5	0.238^{+}	0.418^{*}	0.521^{+}	0.261^{+}	0.187^{**}	0.170^{*}
	(0.0755)	(0.219)	(0.140)	(0.0757)	(0.0879)	(0.0889)
Gener. 2	-0.0430	0.0260	0.0265	-0.0297	0.00659	0.0103
	(0.0552)	(0.152)	(0.117)	(0.0555)	(0.0598)	(0.0574)
Medium aspi.		0.126^{**}				0.102
		(0.0504)				(0.0632)
High aspi.		0.372^{+}				0.299^{+}
		(0.0436)				(0.110)
Gener. 1.5 × Medium aspi.		-0.0959				
		(0.250)				
Gener. 1.5 × High aspi.		-0.244				
		(0.222)				
Gener. 2 × Medium aspi.		-0.0849				
		(0.154)				
Gener. 2 × High aspi.		-0.0754				
		(0.152)				
Medium Exp.			0.179^{+}			0.0564
			(0.0426)			(0.0591)
High Exp.			0.451^{+}			0.212^{**}
			(0.0405)			(0.107)
Gener. 1.5 × Medium Exp.			-0.195			
			(0.156)			
Gener. 1.5 × High Exp.			-0.392^{+}			
			(0.130)			
Gener. 2 × Medium Exp.			-0.0896			
			(0.126)			
Gener. 2 × High Exp.			-0.0347			
			(0.129)			
Asp.< Exp.				-0.0846^{**}	-0.0673	0.0973
				(0.0385)	(0.0417)	(0.0765)
Asp.> Exp.				-0.158^{+}	-0.165^{+}	-0.0948
				(0.0325)	(0.0372)	(0.0743)
Gener. 1.5 × Asp.< Exp.					-0.139	-0.0942
					(0.205)	(0.187)
Gener. 1.5 × Asp.> Exp.					0.212^{**}	0.224^{**}
					(0.107)	(0.106)
Gener. 2 × Asp.< Exp.					-0.142	-0.141
					(0.131)	(0.129)
Gener. 2 × Asp.> Exp.					-0.0859	-0.0791
					(0.100)	(0.0980)
Constant	2.639^{+}	2.533^{+}	2.471^{+}	2.627^{+}	2.669^{+}	2.487^{+}
	(0.447)	(0.451)	(0.445)	(0.447)	(0.444)	(0.439)
Observations	9119	9119	9119	9119	9119	9119
R^2	0.353	0.371	0.381	0.358	0.359	0.384
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$. The results in this table were estimated using Wave III Education Sample weights.

Table 3.7: OLS regression results for Science GPA (4 year average) with school fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
Gener. 1.5	0.176**	0.427**	0.516+	0.201**	0.129	0.109
	(0.0806)	(0.193)	(0.129)	(0.0825)	(0.0904)	(0.0881)
Gener. 2	-0.0495	0.0499	0.00928	-0.0338	0.00145	0.00435
	(0.0622)	(0.126)	(0.102)	(0.0631)	(0.0678)	(0.0655)
Medium aspi.		0.120**				0.133*
		(0.0537)				(0.0727)
High aspi.		0.331+				0.335+
		(0.0416)				(0.111)
Gener. 1.5 × Medium aspi.		-0.159				
		(0.268)				
Gener. 1.5 × High aspi.		-0.318				
		(0.222)				
Gener. 2 × Medium aspi.		-0.0930				
		(0.139)				
Gener. 2 × High aspi.		-0.112				
		(0.125)				
Medium Exp.			0.202+			0.0726
			(0.0414)			(0.0581)
High Exp.			0.430+			0.125
			(0.0416)			(0.111)
Gener. 1.5 × Medium Exp.			-0.320**			
			(0.150)			
Gener. 1.5 × High Exp.			-0.433+			
			(0.147)			
Gener. 2 × Medium Exp.			-0.0568			
			(0.119)			
Gener. 2 × High Exp.			-0.0323			
			(0.110)			
Asp.< Exp.				-0.0738	-0.0614	0.110
				(0.0535)	(0.0586)	(0.0941)
Asp.> Exp.				-0.171+	-0.176+	-0.181**
				(0.0335)	(0.0371)	(0.0751)
Gener. 1.5 × Asp.< Exp.					-0.000162	0.0497
					(0.131)	(0.123)
Gener. 1.5 × Asp.> Exp.					0.191*	0.200*
					(0.112)	(0.113)
Gener. 2 × Asp.< Exp.					-0.145	-0.143
					(0.133)	(0.133)
Gener. 2 × Asp.> Exp.					-0.0832	-0.0794
					(0.0916)	(0.0901)
Constant	2.459+	2.350+	2.273+	2.443+	2.474+	2.291+
	(0.475)	(0.482)	(0.466)	(0.470)	(0.468)	(0.464)
Observations	9091	9091	9091	9091	9091	9091
R^2	0.325	0.337	0.346	0.329	0.330	0.348
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$. The results in this table were estimated using Wave III Education Sample weights.

with fewer TV hours measured one year later. This correlation persists when we control for the current hours spent watching TV in Wave 1. Nevertheless, there is a reduction in the coefficient associated with 1.5 generation and misalignment (Aspirations greater than Expectations). The result suggests that this particular group of students might be dedicating less time for leisure activities and potentially more time to study.

Table 3.8: OLS estimation on the number of hours watching TV per week in Add Health's wave II

	(1) hours tvW2	(2) hours tvW2	(3) hours tvW2
Generation 1.5	-4.044^{+}	-2.151	-1.611
	(1.252)	(1.323)	(1.352)
Generation 2.0	-1.020	-1.488	-1.040
	(0.929)	(0.924)	(0.806)
Asp.< Exp.		0.241	0.351
		(0.811)	(0.760)
Asp.> Exp.		0.831	0.760
		(0.661)	(0.558)
Gener. 1.5 x Asp.< Exp.		0.947	3.252
		(4.431)	(3.850)
Gener. 1.5 x Asp.> Exp.		-6.329^{+}	-5.636^{+}
		(1.665)	(1.505)
Gener. 2 x Asp.< Exp.		-1.976	-3.021*
		(1.921)	(1.718)
Gener. 2 x Asp.> Exp.		1.760	1.370
		(1.889)	(1.828)
Hours tv in t-1			0.354^{+}
			(0.0200)
Observations	8420	8420	8402
R^2	0.130	0.133	0.242
Individual controls	Yes	Yes	Yes
Household controls	Yes	Yes	Yes
Grade	Yes	Yes	Yes
School FE	Yes	Yes	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$. The results in this table were estimated using Wave II Sample weights.

3.5.2 Robustness checks

In the previous sections, we show how the misalignment between aspirations and expectations is a major driving force that explains the over-performance of migrant children. The misalignment is associated with fewer leisure activities for this group of students. In this section, we assess the robustness of our main findings. In addition to the OLS estimation presented in the previous section, the censoring of the GPA between zero and four is addressed in this section. While our measure of GPA is a weighted average by credits, it is continuous over the range zero and four, meaning that students cannot obtain a grade greater than four or smaller than zero. Therefore, to take into account for left- and right-censoring in the dependent variable, we estimate Tobit regressions for our measures of GPA. Tables 3.9 show the results using a Tobit model for high school GPA with zero as the lower limit and four as the upper limit. The size of the coefficients are slightly altered by the use of a Tobit model, nevertheless, the results presented in the tables do not diverge qualitatively from the results reported previously using OLS.

Moreover, we test whether the results are driven by those students who migrated at younger ages. These students have spent more time in the host country; therefore, they are better assimilated and able to achieve similar academic grades when compared to their native counterparts. Cortes (2006) shows that the longer first-generation migrant children live in the USA, the score gap between them and the second-generation immigrant children diminishes. To ensure that our results are not driven by teenagers who migrated at younger ages, we estimate equations (3.1), (3.2) and (3.3) excluding the adolescents who migrated before the age of six. Table 3.10 contains the OLS regression results for overall GPA, excluding this sub-sample of children. When comparing the results from Tables 3.4 and 3.10, we obtain similar estimates. At first glance, 1.5 migrant teens seem to outperform native teens (Column 1).

Table 3.9: Tobit regression results for Overall GPA with school fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
Gener. 1.5	0.166**	0.309	0.474+	0.187+	0.0965	0.0773
	(0.0667)	(0.189)	(0.118)	(0.0677)	(0.0756)	(0.0761)
Gener. 2	-0.0416	-0.0155	0.0162	-0.0285	0.00207	0.00492
	(0.0537)	(0.135)	(0.104)	(0.0538)	(0.0588)	(0.0560)
Medium aspi.		0.153+				0.131**
		(0.0424)				(0.0560)
High aspi.		0.352+				0.292+
		(0.0347)				(0.0856)
Gener. 1.5 × Medium aspi.		-0.0580				
		(0.232)				
Gener. 1.5 × High aspi.		-0.202				
		(0.192)				
Gener. 2 × Medium aspi.		-0.0809				
		(0.135)				
Gener. 2 × High aspi.		-0.0196				
		(0.130)				
Medium Exp.			0.208+			0.0912*
			(0.0342)			(0.0486)
High Exp.			0.441+			0.199**
			(0.0338)			(0.0893)
Gener. 1.5 × Medium Exp.			-0.224*			
			(0.129)			
Gener. 1.5 × High Exp.			-0.438+			
			(0.121)			
Gener. 2 × Medium Exp.			-0.0628			
			(0.113)			
Gener. 2 × High Exp.			-0.0298			
			(0.108)			
Asp.< Exp.				-0.0725*	-0.0565	0.0821
				(0.0386)	(0.0423)	(0.0637)
Asp.> Exp.				-0.149+	-0.161+	-0.120*
				(0.0288)	(0.0322)	(0.0646)
Gener. 1.5 × Asp.< Exp.					-0.0673	-0.0194
					(0.144)	(0.132)
Gener. 1.5 × Asp.> Exp.					0.253+	0.263+
					(0.0959)	(0.0964)
Gener. 2 × Asp.< Exp.					-0.160	-0.158
					(0.103)	(0.102)
Gener. 2 × Asp.> Exp.					-0.0564	-0.0502
					(0.0902)	(0.0869)
Constant	2.959+	2.857+	2.801+	2.949+	2.988+	2.813+
	(0.429)	(0.429)	(0.413)	(0.425)	(0.420)	(0.409)
Observations	9153	9153	9153	9153	9153	9153
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. The goodness of fit measures cannot be displayed after using SVY command in Stata.Standard errors clustered by school are displayed in parentheses. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$. The results in this table were estimated using Wave III Education Sample weights.

However, once we interact with migration generation and misalignment between aspirations and expectations, we find that 1.5 migrants who show higher aspirations than expectations are the ones that score a higher GPA. In Table 3.11, we show the results excluding teenagers who migrated when they were older than 14 years. Once again, we find similar results; nevertheless, it is worth noticing that the coefficient for 1.5 migrant teens in column (1) is marginally smaller when comparing with Table 3.4. When we eliminate the students that migrated at older ages, the difference in the coefficient between 1.5 migrant children and native children diminishes. The above suggests that those students who arrived in the USA after 14 years old might be amplifying the difference in the performance between 1.5 generation migrants and natives.

An additional potential confounding element that could be affecting our results is the adopted definition of children of migrants. We have defined a child of migrants as a teen for which at least one of the biological parents was born outside the USA. Nevertheless, it is also possible that a teen who has one native parent and one migrant parent might have an advantage over the other teens. Children of inter-ethnic parents might differ from children of intra-ethnic couples. For instance, Emonds and van Tubergen (2015) show that the higher human capital and language skills of inter-ethnic couples translate into a better performance of their children. In order to test if our results are not a product of this characteristic among the children of migrants, we reproduce equations (3.1), (3.2) and (3.3) excluding teens who have one native and one migrant parent. The total number of students excluded is 470. The results are presented in Table 3.12. Once again, the results are in line with our previous findings. Since the size of the migrant sample decreases, our standard errors are somewhat larger when compared to Table 3.4. Nevertheless, the results point to conclude that 1.5 generation migrants who outperform at school correspond to those who report high

Table 3.10: OLS regression results for Overall GPA excluding the children who migrated to the USA between 0 and 5 years old

	(1)	(2)	(3)	(4)	(5)	(6)
Gener. 1.5	0.182**	0.421*	0.602+	0.205+	0.113	0.0908
	(0.0760)	(0.224)	(0.139)	(0.0758)	(0.0817)	(0.0839)
Gener. 2	-0.0323	-0.00593	0.0223	-0.0187	0.00849	0.0110
	(0.0533)	(0.135)	(0.105)	(0.0531)	(0.0579)	(0.0551)
Medium aspi.		0.154+				0.125**
		(0.0425)				(0.0566)
High aspi.		0.346+				0.287+
		(0.0347)				(0.0869)
Gener. 1.5 × Medium aspi.		-0.300				
		(0.275)				
Gener. 1.5 × High aspi.		-0.300				
		(0.219)				
Gener. 2 × Medium aspi.		-0.0783				
		(0.135)				
Gener. 2 × High aspi.		-0.0209				
		(0.130)				
Medium Exp.			0.209+			0.0892*
			(0.0346)			(0.0491)
High Exp.			0.438+			0.198**
			(0.0337)			(0.0905)
Gener. 1.5 × Medium Exp.			-0.428+			
			(0.151)			
Gener. 1.5 × High Exp.			-0.551+			
			(0.132)			
Gener. 2 × Medium Exp.			-0.0569			
			(0.114)			
Gener. 2 × High Exp.			-0.0289			
			(0.108)			
Asp.< Exp.				-0.0680*	-0.0532	0.0853
				(0.0388)	(0.0418)	(0.0639)
Asp.> Exp.				-0.154+	-0.161+	-0.118*
				(0.0283)	(0.0317)	(0.0656)
Gener. 1.5 × Asp.< Exp.					-0.144	-0.0413
					(0.193)	(0.184)
Gener. 1.5 × Asp.> Exp.					0.246**	0.259+
					(0.0947)	(0.0987)
Gener. 2 × Asp.< Exp.					-0.152	-0.150
					(0.104)	(0.102)
Gener. 2 × Asp.> Exp.					-0.0514	-0.0458
					(0.0876)	(0.0845)
Constant	2.948+	2.841+	2.773+	2.932+	2.977+	2.800+
	(0.423)	(0.425)	(0.409)	(0.419)	(0.414)	(0.404)
Observations	8915	8915	8915	8915	8915	8915
R^2	0.398	0.416	0.430	0.403	0.404	0.432
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. Significance levels * $p < 0.10$, ** $p < 0.05$, + $p < 0.01$. The results in this table were estimated using Wave III Education Sample weights.

Table 3.11: OLS regression results for Overall GPA excluding the children who migrated to the USA when they were older than 14 years old

	(1)	(2)	(3)	(4)	(5)	(6)
Gener. 1.5	0.161**	0.263	0.437+	0.179**	0.101	0.0774
	(0.0712)	(0.206)	(0.122)	(0.0723)	(0.0787)	(0.0788)
Gener. 2	-0.0352	-0.00523	0.0274	-0.0222	0.00552	0.00787
	(0.0515)	(0.133)	(0.103)	(0.0515)	(0.0561)	(0.0536)
Medium aspi.		0.151+				0.132**
		(0.0425)				(0.0557)
High aspi.		0.348+				0.287+
		(0.0347)				(0.0840)
Gener. 1.5 × Medium aspi.		0.0541				
		(0.245)				
Gener. 1.5 × High aspi.		-0.165				
		(0.212)				
Gener. 2 × Medium aspi.		-0.0789				
		(0.135)				
Gener. 2 × High aspi.		-0.0251				
		(0.129)				
Medium Exp.			0.208+			0.0964**
			(0.0342)			(0.0477)
High Exp.			0.438+			0.202**
			(0.0335)			(0.0878)
Gener. 1.5 × Medium Exp.			-0.180			
			(0.136)			
Gener. 1.5 × High Exp.			-0.404+			
			(0.128)			
Gener. 2 × Medium Exp.			-0.0671			
			(0.112)			
Gener. 2 × High Exp.			-0.0384			
			(0.108)			
Asp.< Exp.				-0.0715*	-0.0554	0.0781
				(0.0384)	(0.0419)	(0.0627)
Asp.> Exp.				-0.150+	-0.160+	-0.118*
				(0.0288)	(0.0320)	(0.0640)
Gener. 1.5 × Asp.< Exp.					-0.104	-0.0563
					(0.172)	(0.160)
Gener. 1.5 × Asp.> Exp.					0.233**	0.248**
					(0.106)	(0.108)
Gener. 2 × Asp.< Exp.					-0.153	-0.153
					(0.104)	(0.103)
Gener. 2 × Asp.> Exp.					-0.0494	-0.0439
					(0.0893)	(0.0861)
Constant	2.992+	2.889+	2.835+	2.984+	3.017+	2.837+
	(0.423)	(0.423)	(0.407)	(0.419)	(0.414)	(0.403)
Observations	9056	9056	9056	9056	9056	9056
R^2	0.397	0.416	0.428	0.402	0.403	0.431
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$. The results in this table were estimated using Wave III Education Sample weights.

ambitions under pessimistic expectations.

Another concern is the temporal order of the measurement of educational outcomes relative to the measurement of aspirations and expectations. It is a concern when using a dependent variable based on all four years of high school transcript data, while aspirations and expectations are measured at Wave 1. For students in later high school grades during Wave 1 (for 12th graders), the aspirations and expectations were measured after the GPA had been partially or almost completely determined. The above could question the casual interpretation of the results presented in previous sections. We test whether our results are sensitive to keeping a strict temporal order between the dependent and independent variables by eliminating the students attending 12th grade during Wave I. The results for the overall transcript GPA are presented in Table 3.13. We find once again a significant and positive interaction between being a 1.5 generation migrant and having higher aspirations than expectations. When comparing with Table 3.4, it is noted that the size of the coefficient of this interaction is larger moving from 0.26 in Table 3.4 to 0.29 in Table 3.13 (Columns 6 and 7). Nevertheless, the results do not dissent in qualitative terms from the results reported previously using the full sample.

Table 3.12: OLS regression results for Overall GPA (4 year average) excluding the children with one migrant and one native parent

	(1)	(2)	(3)	(4)	(5)	(6)
Gener. 1.5	0.175**	0.363*	0.500+	0.204**	0.121	0.113
	(0.0851)	(0.206)	(0.129)	(0.0865)	(0.0925)	(0.0939)
Gener. 2	-0.0284	-0.111	-0.0227	-0.00726	0.00376	0.0190
	(0.0665)	(0.123)	(0.117)	(0.0673)	(0.0735)	(0.0710)
Medium aspi.		0.150+				0.128**
		(0.0422)				(0.0584)
High aspi.		0.348+				0.272+
		(0.0345)				(0.0895)
Gener. 1.5 × Medium aspi.		-0.111				
		(0.237)				
Gener. 1.5 × High aspi.		-0.247				
		(0.205)				
Gener. 2 × Medium aspi.		0.155				
		(0.138)				
Gener. 2 × High aspi.		0.0924				
		(0.131)				
Medium Exp.			0.204+			0.0951*
			(0.0340)			(0.0498)
High Exp.			0.434+			0.218**
			(0.0328)			(0.0920)
Gener. 1.5 × Medium Exp.			-0.249*			
			(0.130)			
Gener. 1.5 × High Exp.			-0.436+			
			(0.128)			
Gener. 2 × Medium Exp.			0.00232			
			(0.134)			
Gener. 2 × High Exp.			0.0715			
			(0.125)			
Asp.< Exp.				-0.0585	-0.0603	0.0665
				(0.0378)	(0.0418)	(0.0646)
Asp.> Exp.				-0.145+	-0.158+	-0.102
				(0.0298)	(0.0319)	(0.0670)
Gener. 1.5 × Asp.< Exp.					-0.0659	-0.0236
					(0.153)	(0.138)
Gener. 1.5 × Asp.> Exp.					0.235**	0.242**
					(0.0983)	(0.0989)
Gener. 2 × Asp.< Exp.					0.0631	0.0602
					(0.159)	(0.149)
Gener. 2 × Asp.> Exp.					-0.0364	-0.0261
					(0.102)	(0.0998)
Constant	3.049+	2.949+	2.897+	3.030+	3.064+	2.907+
	(0.416)	(0.418)	(0.405)	(0.414)	(0.410)	(0.401)
Observations	8683	8683	8683	8683	8683	8683
R^2	0.402	0.421	0.433	0.407	0.408	0.435
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$. The results in this table were estimated using Wave III Education Sample weights.

Table 3.13: OLS regression results for Overall GPA (4 year average) with school fixed effects without students in 12 grade

	(1)	(2)	(3)	(4)	(5)	(6)
Gener. 1.5	0.182^{**}	0.324	0.500^{+}	0.200^{**}	0.0947	0.0707
	(0.0784)	(0.242)	(0.136)	(0.0797)	(0.0884)	(0.0891)
Gener. 2	-0.0347	0.0529	0.0815	-0.0214	0.00814	0.0175
	(0.0557)	(0.142)	(0.104)	(0.0559)	(0.0622)	(0.0591)
Medium aspi.		0.162^{+}				0.143^{**}
		(0.0438)				(0.0594)
High aspi.		0.377^{+}				0.305^{+}
		(0.0369)				(0.0926)
Gener. 1.5 × Medium aspi.		-0.0613				
		(0.276)				
Gener. 1.5 × High aspi.		-0.204				
		(0.234)				
Gener. 2 × Medium aspi.		-0.148				
		(0.145)				
Gener. 2 × High aspi.		-0.0893				
		(0.140)				
Medium Exp.			0.233^{+}			0.109^{**}
			(0.0350)			(0.0505)
High Exp.			0.455^{+}			0.207^{**}
			(0.0379)			(0.0958)
Gener. 1.5 × Medium Exp.			-0.224			
			(0.143)			
Gener. 1.5 × High Exp.			-0.472^{+}			
			(0.132)			
Gener. 2 × Medium Exp.			-0.133			
			(0.116)			
Gener. 2 × High Exp.			-0.0934			
			(0.113)			
Asp.< Exp.				-0.0974^{**}	-0.0808	0.0641
				(0.0472)	(0.0515)	(0.0722)
Asp.> Exp.				-0.142^{+}	-0.156^{+}	-0.118^{*}
				(0.0308)	(0.0340)	(0.0696)
Gener. 1.5 × Asp.< Exp.					-0.109	-0.0661
					(0.199)	(0.200)
Gener. 1.5 × Asp.> Exp.					0.280^{+}	0.290^{+}
					(0.0989)	(0.0996)
Gener. 2 × Asp.< Exp.					-0.194	-0.211
					(0.143)	(0.144)
Gener. 2 × Asp.> Exp.					-0.0445	-0.0482
					(0.0989)	(0.0969)
Constant	2.963^{+}	2.800^{+}	2.763^{+}	2.972^{+}	3.003^{+}	2.743^{+}
	(0.441)	(0.438)	(0.415)	(0.437)	(0.431)	(0.415)
Observations	7707	7707	7707	7707	7707	7707
R^2	0.397	0.417	0.429	0.402	0.403	0.432
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$. The results in this table were estimated using Wave III Education Sample weights.

3.5.3 Omitted variable bias and coefficient stability

We employ Oster (2019) approach to test for the stability of the coefficients of interest that considers the variance explained by the controls. The underlying idea is that if a coefficient is invariant after including the observed controls, hence the omitted variable bias is narrow. Tables 3.17 to 3.20 (in Appendix B) show the stability of the results for overall GPA, Math, English literature, and Science. The tables present a step-wise inclusion of individual, family, and school control variables. Moreover, we present in Table 3.14 the adjusted coefficients for 1.5 generation migrants, misalignment, and the interaction of misalignment and 1.5 generation migrants. The unbiased-adjusted coefficients are based on the assumption that the unobservable determinants explain as much of the variation in the outcome as the observable variables.

Table 3.14 also presents the different calculated δ. These deltas show the degree of importance that the unobservable determinants relative to the observables would need to have for the treatment effect to be zero. Oster's δ values are calculated for a maximum R^2 corresponding to 1.3 times the R^2 of the specification that includes all the controls variables. The results indicate that the degree of selection on unobservables needs to be between 10.8 to 34 times that of the degree of selection on observables so that the omitted variable bias is important enough to make the value of the coefficient associated with the interaction between 1.5 generation migrants and misalignment to be statistically non-significant. In the estimations, we observe that the degree of selection on unobservables would have to be in the reversed direction of the bias to alter the coefficient associated with 1.5 migrant children, as shown by the negative sign of δ_1.

Finally, all δ, taken in absolute values, falls farther from the bound (0 to 1) suggested by Oster (2019). The above suggests that the omitted variable bias does not lead to conclude that our results are statistically invalid.

Table 3.14: Omitted variable bias analysis following Oster (2019)

	(1) Overall GPA	(2) Math GPA	(3) English GPA	(4) Science GPA
A) Uncontrolled Coefficients				
Gener. 1.5	-0.0475	-0.0786	0.0605	-0.0399
Asp.> Exp.	-0.396	-0.363	-0.401	-0.412
Gener. 1.5 x Asp.> Exp.	0.323	0.376	0.244	0.256
r^2	0.038	0.026	0.032	0.031
B) Controlled Coefficients				
Gener. 1.5	0.0789	0.0625	0.17	0.109
Asp.> Exp.	-0.119	-0.0994	0.0948	-0.181
Gener. 1.5 x Asp.> Exp.	0.264	0.328	0.224	0.2
r^2	0.43	0.300	0.384	0.348
C) Bias Adjusted Coefficients				
Gener. 1.5	0.120	0.108	0.205	0.158
Asp.> Exp.	-0.027	-0.012	0.257	-0.104
Gener. 1.5 x Asp.> Exp.	0.244	0.312	0.217	0.181
D) Oster δ				
Oster δ Gener. 1.5	-1.896	-1.348	-4.743	-2.222
Oster δ Asp.> Exp.	1.305	1.148	-0.584	2.379
Oster δ Gener. 1.5 x Asp.> Exp.	13.597	20.803	34.222	10.844

Notes: Columns 1–4 present results from OLS specifications. Part A shows the coefficients of a regression without controls. Part B of the table presents the coefficients after adding the full set of controls that includes aspirations, expectations, age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income and school fixed effects. The last two parts shows the analysis of the potential omitted variable bias as proposed by Oster (2019). In part C, we display the bias-adjusted coefficients assuming that the level of selection on observables is equal to the selection on unobservables (δ= 1) with the highest R^2 value equal to 1.3*R^2 of the specification that includes all the control variables. In the part D, we calculate Oster δ for Gener. 1.5 ; Asp.> Exp and Gener. 1.5 x Asp.> Exp, for a null hypothesis of zero and for a the highest R^2 value equal to 1.3*R^2 of the specification that includes all the control variables. * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$.

3.6 Conclusion

Understanding the educational outcomes of the children of migrants is deemed to be critical for the eventual integration of migrants in western countries. An extensive literature has uncovered an apparent educational advantage of immigrants' children in the USA after controlling for different socio-economic characteristics such as family income and parental education. What explains the over-achievement or super-achievement of the children of migrants in the US? This paper aims to answer this question by studying the gap between educational aspirations and expectations as a potential driving force behind the academic performance of immigrant children. The data used is the National Longitudinal Study of Adolescent to Adult Health (Add Health) collected by the Carolina Population Center. The Add Health study contains detailed information on academic performance, parental information of native and immigrant children in the US and school characteristics. This database follows a group of students born between 1974 and 1983 who studied within the USA's school system between 1994 and 2002. On the one hand, the data confirms that 1.5 migrant generation teens exhibit greater aspirations to achieve high education than their peers do. The results are similar to previous literature that confirms the optimism among children of migrants. For example, Tjaden and Hunkler (2017) and Tjaden and Scharenberg (2017) found that migrant students in Germany and Switzerland are expressing high aspirations to achieve a university degree by choosing the academic track instead of vocational education. On the other hand, migrant teens surveyed in the Add Health study are less optimistic about their chances to achieve those dreams since they report lower expectations to obtain high educational degrees. After controlling for an extensive list of individual, family, and context variables, we document that aspirations, by themselves, are not sufficient to explain the overperformance of migrant children. In contrast, our paper suggests that misalignment between their aspirations and expectations motivates migrant children to increase their efforts to compensate for their perceived disadvantage.

In addition, our paper documents that once various socio-economic and school variables are accounted for, we find no difference between the school performance of second-generation migrants and natives.

To dig deeper into why children of immigrants in the USA perform surprisingly well in school, we explore effort as an underlining mechanism that links motivation with future outcomes. We make use of leisure time as a substitute for studying time. We study whether migrant children spend more or less time watching television compared to their peers. The results show that 1.5 generation migrant teens with misaligned aspirations watch less TV in the subsequent year, suggesting that migrant students who report lower expectations than aspirations might dedicate more time to study to compensate for their perceived disadvantages. This paper suggests that misalignment between expectations and aspirations acts as a driving force for migrant children and is associated with a higher average GPA than their peers.

Needless to say, given the particularities of the American society and its schooling system, the reported positive response of the immigrant children in this study cannot be generalized to every context. We nevertheless think that these results are informative about how immigrant children can display different adjustments in comparison to native pupils. Future research could replicate these results in other contexts or using other surveys.

Acknowledgments

This research uses data from Add Health, a project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this particular study.

3.7 Appendix B

Table 3.15: Description of the variables in Addhealth - Part 1

Variable	Description
College Aspirations	Categories: Low (1 to 3), Medium (4) and High (5). Students were asked to select from a scale of 1 (low) to 5 (high), how much do they want to go to college.
College Expectations	Categories: Low (1 to 3), Medium (4) and High (5). Students were asked to select from a scale of 1 (low) to 5 (high), how likely is it that they will attend college
Age	Age of the student during in Wave I
Male	Dummy variable equal to one if the student is male and zero otherwise.
Body-Mass-Index	Body weight in kg / (height in meters)^2
PPVT	Corresponds to the score of the Peabody Picture Vocabulary Test. It is a standardized test to assess the verbal intelligence of an individual
Self-esteem index	An index constructed using seven questions. The students were asked to agree or disagree with the following questions: 1) Do you have a lot of good qualities? 2) Are you physically fit? 3) Do you have a lot to be proud of? 4) Do you like yourself just the way you are? 5) Do you feel like you are doing everything just about right? 6) Do you feel socially accepted? 7) Do you feel loved and wanted?
Internal Locus	Dichotomous variable equal to one if the student agreed to the following statement: "When you get what you want, it's usually because you worked hard for it".

(continues)

Table 3.16: Description of the variables in Add Health - Part 2

Variable	Description
Ethnicity	Categories: Hispanic, White non-hispanic, asian, black non-hispanic and others
English at Home	Dummy variable equal to one if the family speaks English at home and zero otherwise.
N siblings	Number of siblings living in the household in Wave I
Parent education	Categories: college graduate, high school graduate, less than high school, missing information. We used both the child answers and the mother's reply in order to reduce the number of missing values.
Parental contextual attainment	The share of individuals of the same age category of the parent's origin country who have lower or the same level of education (Feliciano and Lanuza, 2017) as the parent. For the case of native born parents, we used the mother education. Data for the educational distribution in the origin country comes form Barro-Lee dataset.
Family structure	Categories: Both biological parents, at least one step-parent, single parent or other.
Parental expectations	Parents were asked in Wave 1, "how disappointed would you be if [your child] did not graduate from college?" Answers enclosed in three categories: very disappointed, somewhat disappointed, not disappointed; we use not disappointed as the reference category.
Parent involvement index	An index created using the following question: "Which of the things listed on this card have you done with [your mother/adoptive mother/stepmother/foster mother/etc]. in the past 4 weeks". 1. Talked about your school grades or work 2) Worked on a school project 3) Talked about other things you have done in school
Income	Parents were asked the total income before taxes received by the family in 1994. In our descriptive tables, we report total income however, we use the log of income for all regressions.

Table 3.17: Stability of coefficients - OLS regression results for Overall GPA (4 year average) including fixed effects at the school level

	(1) gpal	(2) gpal	(3) gpal	(4) gpal	(5) gpal	(6) gpal	(7) gpal
Gener. 1.5	-0.0475	-0.110	0.145	0.0834	0.0227	0.0331	0.0789
	(0.0979)	(0.0979)	(0.0927)	(0.105)	(0.0956)	(0.0906)	(0.0756)
Gener. 2	0.0881*	-0.0360	0.0897**	0.109*	-0.0258	-0.0340	0.00304
	(0.0522)	(0.0583)	(0.0397)	(0.0568)	(0.0467)	(0.0472)	(0.0543)
Asp.< Exp.	-0.104**	-0.0498	0.0587	-0.110**	0.0769	0.0807	0.0800
	(0.0488)	(0.0489)	(0.0700)	(0.0457)	(0.0692)	(0.0666)	(0.0630)
Asp.> Exp.	-0.396+	-0.258+	-0.121*	-0.340+	-0.101	-0.0948	-0.119*
	(0.0377)	(0.0356)	(0.0720)	(0.0351)	(0.0700)	(0.0689)	(0.0645)
Gener. 1.5 × Asp.< Exp.	-0.0310	-0.0232	-0.0561	0.0414	-0.0562	-0.121	-0.0160
	(0.193)	(0.167)	(0.166)	(0.166)	(0.149)	(0.150)	(0.131)
Gener. 1.5 × Asp.> Exp.	0.323+	0.318+	0.299+	0.330+	0.277+	0.265**	0.264+
	(0.123)	(0.106)	(0.110)	(0.113)	(0.103)	(0.103)	(0.0957)
Gener. 2 × Asp.< Exp.	-0.218	-0.193	-0.192**	-0.144	-0.175*	-0.180*	-0.151
	(0.133)	(0.120)	(0.0930)	(0.130)	(0.0951)	(0.0966)	(0.102)
Gener. 2 × Asp.> Exp.	-0.0397	-0.0184	-0.0389	-0.0143	-0.0323	-0.0374	-0.0437
	(0.0897)	(0.0863)	(0.0836)	(0.0891)	(0.0823)	(0.0815)	(0.0859)
Observations	9153	9153	9153	9153	9153	9153	9153
R^2	0.038	0.176	0.317	0.194	0.347	0.359	0.430
Individual	No	No	Yes	No	Yes	Yes	Yes
Household	No	Yes	No	No	Yes	Yes	Yes
Grade fixed effects	No	No	No	No	No	Yes	Yes
School fixed effects	No	No	No	Yes	No	No	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. The results in this table were estimated using Wave III Education Sample weights. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$.

Table 3.18: Stability of coefficients - OLS regression results for Math (4 year average) including fixed effects at the school level

	(1) math	(2) math	(3) math	(4) math	(5) math	(6) math	(7) math
Gener. 1.5	-0.0786	-0.157	0.118	0.124	-0.0130	-0.00239	0.0625
	(0.114)	(0.118)	(0.101)	(0.122)	(0.111)	(0.106)	(0.0971)
Gener. 2	0.105*	-0.0249	0.139^{+}	0.151**	0.0132	0.00576	0.0351
	(0.0611)	(0.0674)	(0.0468)	(0.0622)	(0.0589)	(0.0580)	(0.0642)
Asp.< Exp.	-0.0471	0.00404	0.122	-0.0524	0.137	0.143*	0.126
	(0.0591)	(0.0569)	(0.0880)	(0.0587)	(0.0852)	(0.0847)	(0.0824)
Asp.> Exp.	-0.363^{+}	-0.234^{+}	-0.128	-0.305^{+}	-0.106	-0.102	-0.0994
	(0.0475)	(0.0446)	(0.0963)	(0.0432)	(0.0953)	(0.0950)	(0.0876)
Gener. 1.5 × Asp.< Exp.	-0.0853	-0.0969	-0.0952	0.0178	-0.111	-0.176	-0.0567
	(0.228)	(0.228)	(0.179)	(0.165)	(0.182)	(0.176)	(0.137)
Gener. 1.5 × Asp.> Exp.	0.376^{+}	0.356^{+}	0.377^{+}	0.385^{+}	0.343^{+}	0.332^{+}	0.328^{+}
	(0.133)	(0.118)	(0.127)	(0.127)	(0.120)	(0.119)	(0.118)
Gener. 2 × Asp.< Exp.	-0.229	-0.201	-0.216*	-0.158	-0.194*	-0.202*	-0.175
	(0.142)	(0.130)	(0.116)	(0.139)	(0.113)	(0.113)	(0.117)
Gener. 2 × Asp.> Exp.	-0.110	-0.0977	-0.1000	-0.0341	-0.0966	-0.0994	-0.0676
	(0.109)	(0.109)	(0.106)	(0.109)	(0.106)	(0.105)	(0.104)
Observations	9124	9124	9124	9124	9124	9124	9124
R^2	0.026	0.114	0.203	0.158	0.224	0.232	0.300
Individual	No	No	Yes	No	Yes	Yes	Yes
Household	No	Yes	No	No	Yes	Yes	Yes
Grade fixed effects	No	No	No	No	No	Yes	Yes
School fixed effects	No	No	No	Yes	No	No	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. The results in this table were estimated using Wave III Education Sample weights. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$.

Table 3.19: Stability of coefficients - OLS regression results for English Literaure (4 year average) including fixed effects at the school level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	engl	engl	engl	engl	engl	engl	engl
Gener. 1.5	0.0605	-0.00262	0.215**	0.138	0.109	0.119	0.170*
	(0.102)	(0.103)	(0.108)	(0.112)	(0.109)	(0.104)	(0.0889)
Gener. 2	0.0922	-0.0186	0.0727	0.0865	-0.0245	-0.0310	0.0103
	(0.0619)	(0.0617)	(0.0516)	(0.0613)	(0.0517)	(0.0517)	(0.0574)
Asp.< Exp.	-0.113**	-0.0575	0.0736	-0.124+	0.0897	0.0911	0.0973
	(0.0509)	(0.0495)	(0.0828)	(0.0444)	(0.0815)	(0.0800)	(0.0765)
Asp.> Exp.	-0.401+	-0.265+	-0.0840	-0.358+	-0.0658	-0.0583	-0.0948
	(0.0426)	(0.0426)	(0.0807)	(0.0405)	(0.0793)	(0.0789)	(0.0743)
Gener. 1.5 $\times$ Asp.< Exp.	-0.130	-0.116	-0.148	-0.0638	-0.147	-0.211	-0.0942
	(0.227)	(0.208)	(0.204)	(0.225)	(0.192)	(0.197)	(0.187)
Gener. 1.5 $\times$ Asp.> Exp.	0.244*	0.249**	0.233*	0.268**	0.219*	0.211*	0.224**
	(0.131)	(0.120)	(0.122)	(0.125)	(0.119)	(0.116)	(0.106)
Gener. 2 $\times$ Asp.< Exp.	-0.206	-0.175	-0.175	-0.148	-0.150	-0.158	-0.141
	(0.146)	(0.138)	(0.113)	(0.155)	(0.115)	(0.118)	(0.129)
Gener. 2 $\times$ Asp.> Exp.	-0.0644	-0.0453	-0.0752	-0.0440	-0.0682	-0.0736	-0.0791
	(0.0943)	(0.0929)	(0.0926)	(0.0962)	(0.0917)	(0.0911)	(0.0980)
Observations	9119	9119	9119	9119	9119	9119	9119
R^2	0.032	0.140	0.278	0.157	0.302	0.311	0.384
Individual	No	No	Yes	No	Yes	Yes	Yes
Household	No	Yes	No	No	Yes	Yes	Yes
Grade fixed effects	No	No	No	No	No	Yes	Yes
School fixed effects	No	No	No	Yes	No	No	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. The results in this table were estimated using Wave III Education Sample weights. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$.

Table 3.20: Stability of coefficients - OLS regression results for Science (4 year average) including fixed effects at the school level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	scie	scie	scie	scie	scie	scie	scie
Gener. 1.5	-0.0399	-0.0950	0.153*	0.0782	0.0309	0.0411	0.109
	(0.110)	(0.115)	(0.0902)	(0.129)	(0.0968)	(0.0923)	(0.0881)
Gener. 2	0.104	-0.0218	0.108**	0.0877	-0.0115	-0.0207	0.00435
	(0.0639)	(0.0740)	(0.0521)	(0.0632)	(0.0595)	(0.0592)	(0.0655)
Asp.< Exp.	-0.107*	-0.0510	0.107	-0.116**	0.122	0.126	0.110
	(0.0611)	(0.0594)	(0.100)	(0.0572)	(0.0997)	(0.0971)	(0.0941)
Asp.> Exp.	-0.412+	-0.269+	-0.187**	-0.355+	-0.167**	-0.160**	-0.181**
	(0.0409)	(0.0385)	(0.0821)	(0.0386)	(0.0810)	(0.0794)	(0.0751)
Gener. 1.5 × Asp.< Exp.	0.0526	0.0566	0.0285	0.110	0.0217	-0.0389	0.0497
	(0.193)	(0.189)	(0.155)	(0.163)	(0.149)	(0.146)	(0.123)
Gener. 1.5 × Asp.> Exp.	0.256*	0.245**	0.240*	0.267**	0.212*	0.201*	0.200*
	(0.130)	(0.117)	(0.122)	(0.124)	(0.117)	(0.116)	(0.113)
Gener. 2 × Asp.< Exp.	-0.249	-0.203	-0.222*	-0.150	-0.191	-0.192	-0.143
	(0.158)	(0.140)	(0.127)	(0.162)	(0.122)	(0.124)	(0.133)
Gener. 2 × Asp.> Exp.	-0.0859	-0.0600	-0.0847	-0.0531	-0.0739	-0.0786	-0.0794
	(0.0955)	(0.0916)	(0.0921)	(0.0910)	(0.0897)	(0.0889)	(0.0901)
Observations	9091	9091	9091	9091	9091	9091	9091
R^2	0.031	0.137	0.249	0.162	0.272	0.280	0.348
Individual	No	No	Yes	No	Yes	Yes	Yes
Household	No	Yes	No	No	Yes	Yes	Yes
Grade fixed effects	No	No	No	No	No	Yes	Yes
School fixed effects	No	No	No	Yes	No	No	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. The results in this table were estimated using Wave III Education Sample weights. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$.

Table 3.21: OLS regression results for gpal GPA (4 year average) with community fixed effects

	(1) gpal	(2) gpal	(3) gpal	(4) gpal	(5) gpal	(6) gpal
Gener. 1.5	0.191^{+}	0.387^{**}	0.516^{+}	0.212^{+}	0.122	0.102
	(0.0682)	(0.191)	(0.119)	(0.0688)	(0.0769)	(0.0780)
Gener. 2	-0.0290	0.0100	0.0283	-0.0153	0.0130	0.0137
	(0.0520)	(0.132)	(0.101)	(0.0522)	(0.0563)	(0.0533)
Medium aspi.		0.157^{+}				0.124^{**}
		(0.0423)				(0.0566)
High aspi.		0.351^{+}				0.275^{+}
		(0.0339)				(0.0863)
Gener. 1.5 × Medium aspi.		-0.119				
		(0.232)				
Gener. 1.5 × High aspi.		-0.260				
		(0.191)				
Gener. 2 × Medium aspi.		-0.0956				
		(0.133)				
Gener. 2 × High aspi.		-0.0357				
		(0.128)				
Medium Exp.			0.214^{+}			0.101^{**}
			(0.0341)			(0.0488)
High Exp.			0.440^{+}			0.212^{**}
			(0.0329)			(0.0897)
Gener. 1.5 × Medium Exp.			-0.251^{*}			
			(0.128)			
Gener. 1.5 × High Exp.			-0.463^{+}			
			(0.118)			
Gener. 2 × Medium Exp.			-0.0690			
			(0.111)			
Gener. 2 × High Exp.			-0.0286			
			(0.107)			
Asp.< Exp.				-0.0657^{*}	-0.0528	0.0762
				(0.0384)	(0.0421)	(0.0635)
Asp.> Exp.				-0.147^{+}	-0.158^{+}	-0.109^{*}
				(0.0285)	(0.0320)	(0.0651)
Gener. 1.5 × Asp.< Exp.					-0.0555	-0.0148
					(0.142)	(0.131)
Gener. 1.5 × Asp.> Exp.					0.247^{+}	0.257^{+}
					(0.0930)	(0.0934)
Gener. 2 × Asp.< Exp.					-0.129	-0.129
					(0.101)	(0.0996)
Gener. 2 × Asp.> Exp.					-0.0576	-0.0501
					(0.0853)	(0.0821)
Constant	3.032^{+}	2.901^{+}	2.826^{+}	3.038^{+}	3.065^{+}	2.857^{+}
	(0.408)	(0.411)	(0.395)	(0.404)	(0.400)	(0.391)
Observations	9153	9153	9153	9153	9153	9153
R^2	0.385	0.404	0.417	0.390	0.391	0.419
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. The results in this table were estimated using Wave III Education Sample weights.. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$.

Table 3.22: OLS regression results for math GPA (4 year average) with community fixed effects

	(1) math	(2) math	(3) math	(4) math	(5) math	(6) math
Gener. 1.5	0.205**	0.367	0.511+	0.225**	0.110	0.0947
	(0.0917)	(0.223)	(0.147)	(0.0921)	(0.0990)	(0.101)
Gener. 2	-0.00904	0.0915	-0.0136	0.00264	0.0475	0.0487
	(0.0622)	(0.133)	(0.105)	(0.0619)	(0.0659)	(0.0637)
Medium aspi.		0.124+				0.0995
		(0.0458)				(0.0677)
High aspi.		0.290+				0.249**
		(0.0417)				(0.109)
Gener. 1.5 × Medium aspi.		-0.0243				
		(0.266)				
Gener. 1.5 × High aspi.		-0.226				
		(0.214)				
Gener. 2 × Medium aspi.		-0.212				
		(0.148)				
Gener. 2 × High aspi.		-0.0951				
		(0.136)				
Medium Exp.			0.147+			0.0549
			(0.0429)			(0.0603)
High Exp.			0.377+			0.163
			(0.0390)			(0.114)
Gener. 1.5 × Medium Exp.			-0.117			
			(0.157)			
Gener. 1.5 × High Exp.			-0.509+			
			(0.138)			
Gener. 2 × Medium Exp.			0.0257			
			(0.129)			
Gener. 2 × High Exp.			0.0376			
			(0.114)			
Asp.< Exp.				-0.0174	-0.000518	0.129
				(0.0505)	(0.0561)	(0.0832)
Asp.> Exp.				-0.133+	-0.144+	-0.100
				(0.0377)	(0.0421)	(0.0883)
Gener. 1.5 × Asp.< Exp.					-0.0986	-0.0646
					(0.145)	(0.141)
Gener. 1.5 × Asp.> Exp.					0.315+	0.324+
					(0.113)	(0.115)
Gener. 2 × Asp.< Exp.					-0.158	-0.158
					(0.113)	(0.114)
Gener. 2 × Asp.> Exp.					-0.110	-0.102
					(0.104)	(0.101)
Constant	3.016+	2.909+	2.857+	3.012+	3.047+	2.881+
	(0.484)	(0.487)	(0.471)	(0.477)	(0.472)	(0.467)
Observations	9124	9124	9124	9124	9124	9124
R^2	0.267	0.277	0.286	0.270	0.271	0.287
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. The results in this table were estimated using Wave III Education Sample weights. Significance levels * $p < 0.10$, ** $p < 0.05$, + $p < 0.01$.

Table 3.23: OLS regression results for engl GPA (4 year average) with community fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
	engl	engl	engl	engl	engl	engl
Gener. 1.5	0.261^{+}	0.491^{**}	0.557^{+}	0.285^{+}	0.211^{**}	0.193^{**}
	(0.0764)	(0.220)	(0.142)	(0.0764)	(0.0888)	(0.0904)
Gener. 2	-0.0262	0.0570	0.0354	-0.0119	0.0238	0.0250
	(0.0547)	(0.150)	(0.116)	(0.0550)	(0.0588)	(0.0562)
Medium aspi.		0.129^{**}				0.0891
		(0.0501)				(0.0637)
High aspi.		0.372^{+}				0.275^{**}
		(0.0432)				(0.111)
Gener. 1.5 × Medium aspi.		-0.134				
		(0.250)				
Gener. 1.5 × High aspi.		-0.301				
		(0.221)				
Gener. 2 × Medium aspi.		-0.111				
		(0.152)				
Gener. 2 × High aspi.		-0.0906				
		(0.151)				
Medium Exp.			0.186^{+}			0.0706
			(0.0425)			(0.0597)
High Exp.			0.454^{+}			0.235^{**}
			(0.0402)			(0.109)
Gener. 1.5 × Medium Exp.			-0.211			
			(0.156)			
Gener. 1.5 × High Exp.			-0.415^{+}			
			(0.130)			
Gener. 2 × Medium Exp.			-0.0837			
			(0.126)			
Gener. 2 × High Exp.			-0.0252			
			(0.128)			
Asp.< Exp.				-0.0780^{**}	-0.0636	0.0882
				(0.0391)	(0.0421)	(0.0766)
Asp.> Exp.				-0.159^{+}	-0.165^{+}	-0.0812
				(0.0325)	(0.0374)	(0.0754)
Gener. 1.5 × Asp.< Exp.					-0.120	-0.0816
					(0.204)	(0.187)
Gener. 1.5 × Asp.> Exp.					0.206^{*}	0.217^{**}
					(0.104)	(0.103)
Gener. 2 × Asp.< Exp.					-0.116	-0.115
					(0.127)	(0.126)
Gener. 2 × Asp.> Exp.					-0.0910	-0.0826
					(0.0961)	(0.0942)
Constant	2.898^{+}	2.760^{+}	2.683^{+}	2.903^{+}	2.933^{+}	2.714^{+}
	(0.443)	(0.448)	(0.441)	(0.441)	(0.439)	(0.437)
Observations	9119	9119	9119	9119	9119	9119
R^2	0.340	0.358	0.368	0.345	0.346	0.371
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

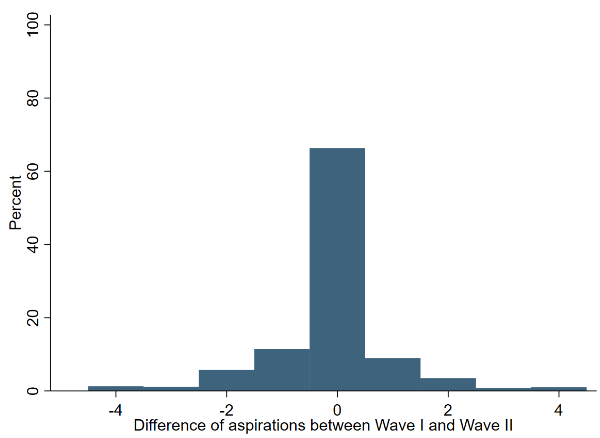
Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$. The results in this table were estimated using Wave III Education Sample weights.

Table 3.24: OLS regression results for scie GPA (4 year average) with community fixed effects

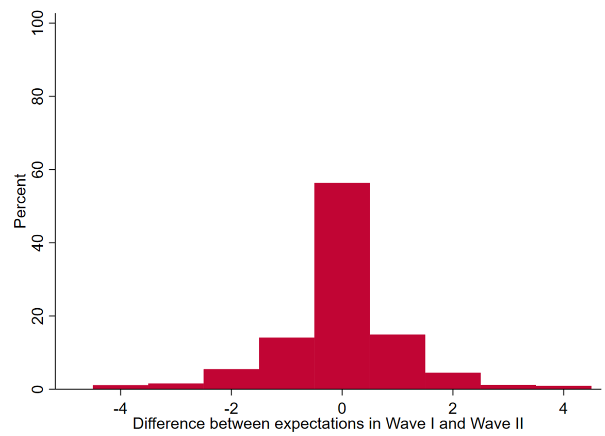
	(1)	(2)	(3)	(4)	(5)	(6)
	scie	scie	scie	scie	scie	scie
Gener. 1.5	0.203**	0.495**	0.557+	0.229+	0.161*	0.141
	(0.0807)	(0.198)	(0.130)	(0.0823)	(0.0906)	(0.0888)
Gener. 2	-0.0319	0.0871	0.0294	-0.0155	0.0187	0.0200
	(0.0620)	(0.126)	(0.102)	(0.0628)	(0.0673)	(0.0650)
Medium aspi.		0.120**				0.117
		(0.0538)				(0.0738)
High aspi.		0.326+				0.311+
		(0.0415)				(0.113)
Gener. 1.5 × Medium aspi.		-0.208				
		(0.271)				
Gener. 1.5 × High aspi.		-0.363				
		(0.225)				
Gener. 2 × Medium aspi.		-0.126				
		(0.140)				
Gener. 2 × High aspi.		-0.132				
		(0.126)				
Medium Exp.			0.208+			0.0829
			(0.0408)			(0.0586)
High Exp.			0.429+			0.143
			(0.0410)			(0.112)
Gener. 1.5 × Medium Exp.			-0.349**			
			(0.150)			
Gener. 1.5 × High Exp.			-0.451+			
			(0.145)			
Gener. 2 × Medium Exp.			-0.0694			
			(0.120)			
Gener. 2 × High Exp.			-0.0324			
			(0.111)			
Asp.< Exp.				-0.0619	-0.0513	0.108
				(0.0537)	(0.0588)	(0.0945)
Asp.> Exp.				-0.169+	-0.174+	-0.166**
				(0.0335)	(0.0373)	(0.0760)
Gener. 1.5 × Asp.< Exp.					-0.00583	0.0383
					(0.130)	(0.123)
Gener. 1.5 × Asp.> Exp.					0.177	0.185*
					(0.109)	(0.110)
Gener. 2 × Asp.< Exp.					-0.122	-0.120
					(0.129)	(0.130)
Gener. 2 × Asp.> Exp.					-0.0858	-0.0810
					(0.0891)	(0.0875)
Constant	2.246+	2.114+	2.024+	2.243+	2.262+	2.060+
	(0.467)	(0.477)	(0.462)	(0.463)	(0.461)	(0.461)
Observations	9091	9091	9091	9091	9091	9091
R^2	0.315	0.327	0.337	0.319	0.320	0.338
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Household controls	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Control variables include age, gender, BMI, PPVT, self-esteem index, internal locus of control, ethnicity, English at home, number of siblings, parent education, parental contextual attainment, family structure, parental expectations for higher education, parent involvement index, household income. Standard errors clustered by school are displayed in parentheses. Significance levels * $p < 0.10$, ** $p < 0.05$, $^{+}$ $p < 0.01$. The results in this table were estimated using Wave III Education Sample weights.

Figure 3.1: Changes in aspirations and expectations to attend college between Wave I and wave II



(a) Expectations to go to college



(b) Aspirations to go to college

Notes: own calculations based on Add Health data. The figure represents the difference between the answers given by the respondent in Wave I versus Wave II.

CHAPTER 4

THE ROLE OF FINANCIAL AID ON FOREIGN EDUCATION: EVIDENCE FROM COLOMBIAN GRADUATES

4.1 Introduction[1]

In today's globalized world, more young adults are seeking to study abroad in order to improve their careers and gain international experience. According to Beine et al. (2014), from 1975 until 2008, the number of international students in OECD countries multiplied by 4. Recently, OECD's International Migration Outlook 2020 reported that in 2018, 1.5 million visas were granted to students pursuing tertiary education in OECD countries. In total, nearly 3.9 million international students were enrolled in OECD's tertiary education institutions by 2018 (OECD, 2020b).[2] While international students are considered as temporary migrants, the magnitude of their inflows is comparable to the inflow of migrants who travel for other motives such as family reunification (1.6 million in 2018) and humanitarian reasons (0.501 million in 2018[3]).

Most international students move temporarily to a foreign country to pursue higher education. Nevertheless, they not only acquire high-quality academic credentials, but they also gain intercultural competence (Salisbury et al., 2013), adaptability, flexibility, and language skills that can result in better and brighter careers (Sorrenti, 2017). These skills are

[1] Ana Cecilia Montes-Viñas gratefully acknowledges the research support from Colfuturo and Jerónimo Castro for providing access to the administrative records.

[2] In contrast, 2.3 million were enrolled in 2006 (International Migration Outlook, 2009).

[3] The statistics reported by OECD's International Migration Outlook 2020 exclude the humanitarian migration flows to Colombia and Turkey in 2018.

desirable to multinational companies (Fielden et al., 2007). The relevance of high-quality education in today's labor market, summed with the cultural experiences acquired while abroad, are among the factors behind the decision of studying abroad expressed by the students in different surveys (Teichler, 2004; Thissen and Ederveen, 2006). Moreover, literature using quasi-experimental methods confirms that studying abroad improves the employability of students once they finish their studies. (Di Pietro, 2015; Parey and Waldinger, 2011; d'Hombres and Schnepf, 2021).[4]

This educational exchange is not only beneficial to students. International students also contribute to the host societies through the payment of fees, cultural transfer, knowledge creation, and innovation (Stuen et al., 2012; Hunt, 2011). Furthermore, the countries from which the international students originate could also gain in this exchange of talent.[5] International students could transfer the knowledge learned abroad after they return to the country of origin (Spilimbergo, 2009). Moreover, the "brain gain" hypothesis suggests that high skilled emigration can potentially increase the overall stock of human capital in the origin country (Mountford, 1997; Stark et al., 1997, 1998; Vidal, 1998). Suppose the returns to education abroad are higher in the destination than in the home country. In that case, the possibility of future migration can create incentives for individuals to improve their education level at their countries of origin.[6]

Whereas international education could bring benefits to young adults, it is a costly en-

[4]While most studies focus on short-term studying abroad experiences such as the ERASMUS program, these studies shed light about the causal effect of studying abroad on employment. See Di Pietro (2019) for a comprehensive literature review.

[5]Although studying abroad increases the chances of living abroad (Oosterbeek and Webbink, 2011; Parey and Waldinger, 2011; Di Pietro, 2012), many students do in fact return to their country of origin. More recent estimates suggest that stay rates in the host country are low. On average, 25% of international students shift from student visa to a different migration status in the host country (OECD, 2011)

[6]A large body of empirical literature support the "brain gain" effect Faini (2006); Beine et al. (2008, 2010b); Gibson and McKenzie (2011); Batista et al. (2012)

deavor.[7] Financial and liquidity constraints are among the factors that prevent students from studying abroad. The role of financial factors on the decision to study overseas has not been widely studied (Whatley, 2017). Some authors have analyzed the role of tuition fees as a determinant for pursuing a degree overseas. Beine et al. (2014), for example, found no significant effects of tuition fees on the inflows of international students. In contrast, Beine et al. (2018) find that university fees reduced the influx of students at the university level after considering the endogeneity of tuition fees. Moreover, Beine et al. (2018) shows that high living costs in the university location decrease the flows of students to that university. I contribute to the literature related to the determinants of studying abroad by exploring financial aid. This paper examines Colfuturo's financial assistance. This program aims to foster higher education degrees abroad for Colombian citizens. On the one hand, students who often apply for financial aid experience financial difficulties. Therefore, less likely to successfully complete an international education (especially in partial scholarships or grants that do not cover the total migration costs). On the other hand, these students might also display great interest in the overseas experience, pushing them to obtain a degree with or without the fund; in this case, the financial aid program might not be relevant.

To my knowledge, there are very few studies that analyze financial aid and its impact on studying abroad. Whatley (2017) analyses the association between study grants and student loans on the probability of studying abroad among students in the university system of Georgia, USA. In this paper, the author shows that student loans decrease the likelihood

[7] Other authors argue that the effect of foreign education is somewhat ambiguous. Studying abroad can also negatively affect labor outcomes if the stay abroad limits the acquisition of human capital specific to the origin country. In addition, if the human capital obtained abroad is not portable or transferable to the "home" context, the individual might face an occupational downgrade (Dustmann and Glitz, 2011). Furthermore, international migration can also affect the creation and maintenance of social networks in their home country affecting the type of jobs after the return. For students from a disadvantaged socioeconomic background, the effect of studying abroad can also be ambiguous if the role of social networks is an essential protagonist in the "home" or origin country's labor market.

of studying abroad while study grants improve it. The above result suggests that students in the USA who take study loans to afford higher education might have significant financial constraints that prevent them from investing additional resources into studying abroad. On the contrary, the students who obtain study grants were more likely to travel overseas to pursue a higher education degree. One limitation of this study is that it cannot consider the students' unobserved characteristics. Students who typically obtain study grants are not a random sample of students; they are high-achieving students with excellent grades. Moreover, students who receive study grants tend to self-select by participating in the selection process of the grant. To mimic a random assignment of the financial aid, Whatley (2019) exploits the gradual implementation the merit-based aids in the USA. The author employs the difference-in-difference approach to study the relationship between state merit-aid programs and the acquisition of foreign education at the bachelor level. Whatley (2019) concluded that merit aid decreases by 0.5 percentage points the participation rate in study abroad among students in the US.

This chapter contributes to the literature in various aspects. Unlike other studies, this paper uses a sample of students who applied for Colfuturo's financial aid program to alleviate potential selection bias. Those who are confident in their ability or highly motivated in pursuing higher education are the most likely to apply for a merit-based study grant. These unobserved determinants of grant application are difficult to account for when a simple comparison between those who obtain a grant and those who did not is made. If these unobserved factors also determine the outcome of interest, then the coefficient would be biased. For example, let's say that those who are the most motivated are the ones who apply and obtain the grant are compared to students who did not obtained it. Then the effect of financial aid on study abroad would be biased upwards when this determinant is not accounted for. In the context analyzed in this chapter, all the students in the sample displayed curiosity for studying abroad, applied to the selection process, and fulfill the requirements for obtaining the grant. The students are selected using a score based on four criteria related to their academic performance, the quality of the university abroad, and an essay. [8] I exploit this characteristic of the selection process to implement a Regression Discontinuity Design (RDD) to uncover the causal effect of receiving the scholarship-loan on the probability of completing studies abroad. Given Colfuturo's selection process, the assignment of the fund is distributed quasi-randomly for the applicants around the vicinity of the cutoff point.

The closest study to the present paper is Oosterbeek and Webbink (2011). The authors use a total sample of 837 applicants to a scholarship program in the Netherlands ("program for the talented") to study the probability of studying abroad and the returns of international students using an RDD. The present study improves previous estimates by analyzing a context in which nearly two thousand individual applications were sent each year to the call between 2011 and 2015. Moreover, while most literature has analyzed studying abroad at the bachelor level or short-term stays such as the ERASMUS program, this chapter adds to the literature by analyzing the study abroad experience at the post-graduate level.

In addition, this chapter contributes to an extensive literature that studies the effect of grants and scholarships on higher education by analyzing this particular conditional grant. The impact of financial aid on higher education has been widely studied when the education is obtained internally (Van Der Klaauw, 2002; Bound and Turner, 2002; Dynarski, 2000, 2003b, 2005; Cornwell et al., 2006; Kane, 2007; Dynarski, 2008; Deming and Dynarski, 2010; Scott-Clayton, 2011). For example, Van Der Klaauw (2002) evaluates the effects of university financial aid awards on college enrollment. The author estimates an elasticity of

[8] see https://www.colfuturo.org/english

college enrollment with respect to the grant of 0.89. Based on a merit scholarship, Dynarski (2000) finds that 1000 US dollars in aid given through Georgia's HOPE Scholarship boosts college attendance rates by 3.7 to 4.2 percentage points. Using also HOPE scholarships, Cornwell et al. (2006) finds an increase in freshmen enrollment by 5.9%. Moreover, using the elimination of the Social Security Student Benefit Program in 1982 in the US, Dynarski (2003) estimates that 1,000 USD in financial aid increments by 3.6 percentage points the probability of attending college.

The size and the relevance of the effect of financial aid on higher education would depend on the type of financial aid. For example, in merit-based aids, financial support has substantial positive effects on both attendance and completion. For the case of completion, the estimated impact of receiving a grant range from 9 to 15 percentage points(Dynarski, 2003b, 2008; Scott-Clayton, 2011). In contrast, Dynarski (2003a) finds evidence of a small effect of study loans on college attendance using the Higher Education Amendments of 1992 in the USA. Similarly, a study by Reyes (1995) discovers that 1,000 USD in loan subsidies increments university attendance by only 1.5 percentage points. The effect of grants versus loans on higher education is an empirical question. Colfuturo's financial aid combines different elements. It is a merit-based aid since it is allocated based on academic performance and does not rest on the applicants' financial needs. Moreover, it is a partial scholarship that can turn into a study loan if the recipient chooses not to return to Colombia after completing the studies. Overall the financial aid offered by Colfuturo display different features that could offset the positive effect of the aid on studying abroad and may result in a not relevant program. To summarize, this paper aims to answer the questions: does Colfuturo's Scholarship - Loan program increase or not change the probability of completing a master's degree abroad?

I use data provided by the Colfuturo foundation for 5478 applicants to the program in 2011, 2012, 2014, and 2015 to study for a master's degree. I combined these record with information gathered from online sources. Using both Local Polynomial Regression and Two-Stage Least Squares (2SLS), I find that the effect of being a recipient of the Colfuturo's financial help increased the probability of studying abroad of Colombian graduates by 30 percentage points approximately. I obtain similar results by including control variables, clustering the standard errors, different polynomial degrees, and different bandwidths around the cutoff point.

The paper is organized as follows. Section 2 provides background about Colfuturo and the Scholarship-Loan program, the selection process, the requirements for the scholarship, and some stylized facts. Section 3 explains the data collection process and presents the description of the sample. Section 4 describes the empirical strategy. Section 5 describes the baseline results as well as the results from different robustness checks. Finally, Section 6 presents a discussion of the results in line with related literature and conclusions related to the external validity of the study and caveats to the analysis.

4.2 Colfuturo Scholarship - Loan program

The Colfuturo Scholarship - Loan program was initiated in 1992 with the intention of training highly qualified human capital. The lack of this kind of human capital in Colombia's productive apparatus, research institutions, and even in the public sector is one of the leading causes of the low rates of efficiency and labor productivity in the country (DNP, 2015). The initiative emerged from the first lady's office with funds provided by companies in the private sector (56%) and capital provided by the public sector (44%). Since 2007, the program added the sponsorship of the Administrative Department of Science, Technology, and Innovation

(Colciencias), the National Education Ministry, and the Colombian Institute of Educational Credit and Technical Studies Abroad (Icetex). The Colfuturo foundation administers the funds. The program provides Colombian students the possibility to conduct their studies abroad by offering financial aid. Students holding an undergraduate degree can apply to receive financial assistance up to 50,000 US dollars to finance their post-graduate studies in any country in the world. For master's degrees that last only one year, the total amount is 25,000 US dollars. The scholarship component is approved when the student returns to Colombia after completing of his/her studies and staying in the country between 3 to 5 years. If the student completed their studies and returned to Colombia, a percentage of the loan is "Forgiven" (Scholarship component). If the student does not return to Colombia or fulfils the degree, they must pay back the total amount of money received.

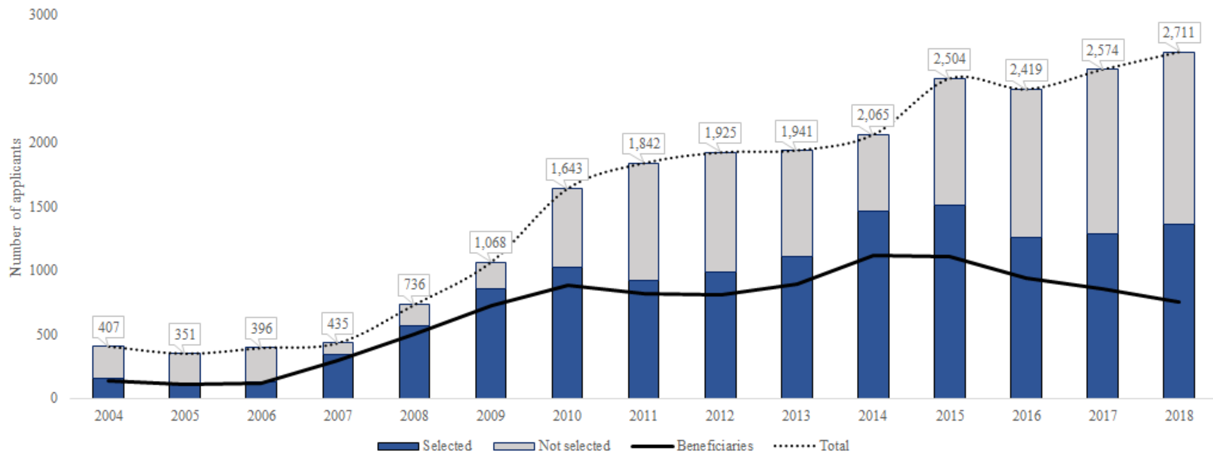
Students are selected based on four criteria: First, the average grade point average (GPA) obtained in the bachelor's studies (42%); Second, the relative ranking in the bachelor's studies (15%) with respect with their graduating cohort; Third, the quality of the post-graduate program (35%) using Colfuturo's ranking which is based on different international rankings and experts opinions; and forth, an essay that states the motivations of their studies (5%) scored by external reviewers [9]. The selection process is summarized as follows: First, the team in Colfuturo receives the information and calculates the score equal to the weighted average of the four criteria. Second, the list is divided by study areas, and a first committee revises the score. This committee is specific for each study area and is composed of four to six professionals and/or professors specialized in each study area and who also obtained a post-graduate degree abroad. The students are ranked from the highest to the lowest based on this score. Third, the ordered list is revised by a second committee made up of 8

[9] For the 2020 application call, there were 21 essay reviewers in total.

professionals in the subject. This second committee selects the students ranked above the cutoff point. The cutoff is defined based on the number of scholarships available for each year and study area. Finally, the list is revised by Colfuturo's Board of Directors, who do the final revision after adding or discarding candidates depending on the final budget for each of the study areas. The average selection rate is 54% overall for both Master's and Doctorate Degree applicants between 1992 to 2018 [10]. Figure 4.1 shows the composition of the selected and non-selected applicants over the years. The figure shows that there is not a clear time trend with regards to the selection rate throughout the years after 2010. The total amount of scholarships to grant depends on year by year of the total capital donated by the sponsors. The government, one of the biggest donors today, plans the allocation of resources to the program for each year through the National Council for Economic and Social Policy (CONPES). For example, in 2015, the central government redefined the distribution of resources to the program until 2025 (DNP, 2015).

[10] The calculation does not include all the applications, only those who delivered all the documents and fulfilled the requirements.

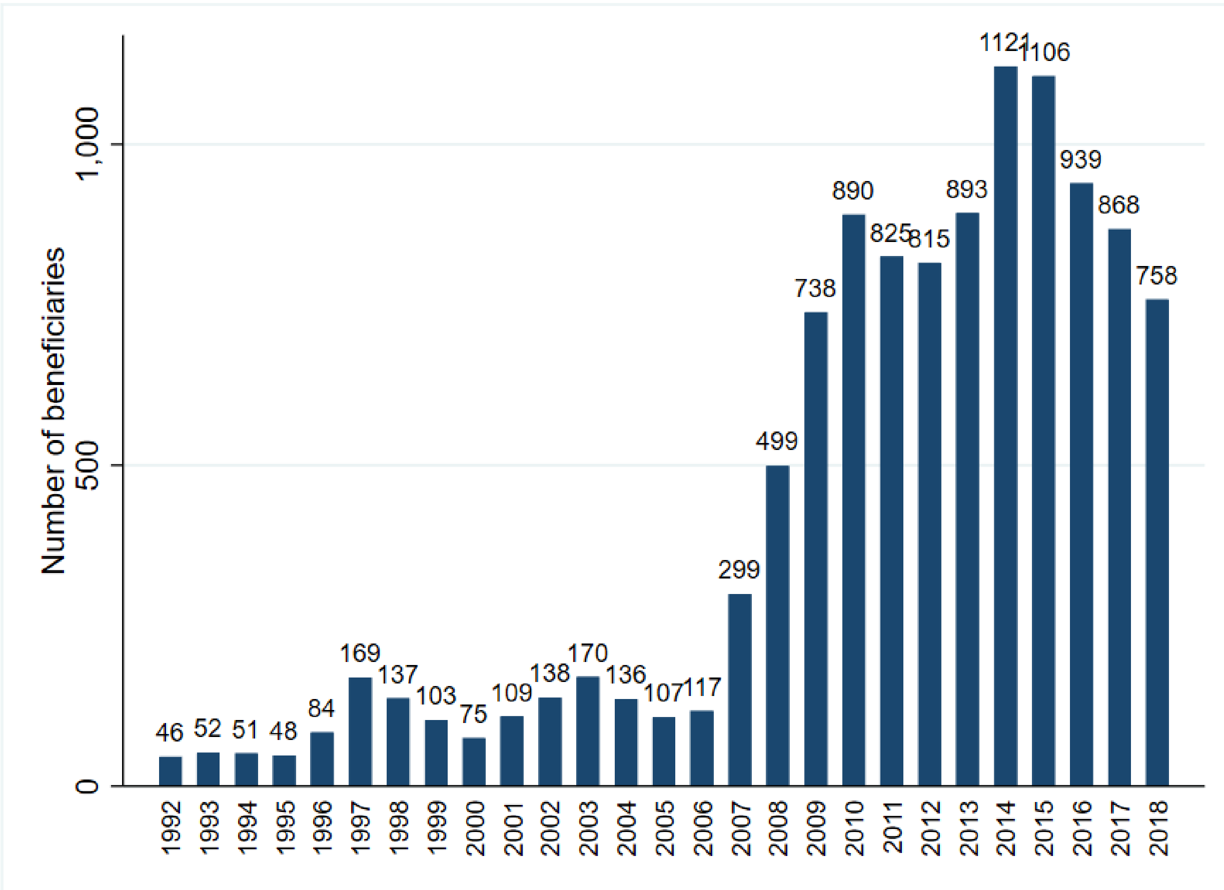
Figure 4.1: The number of total applicants to the Colfuturo scholarship/loan between 2004 to 2018



Source: Colfuturo (2018) Annual Report

The program has disbursed almost 600 million dollars to finance the studies of more than 17 thousand students in nearly 57 countries in the world from 1992 until 2020. The six most selected destinations by Colfuturo's recipients are the United States of America (26%), followed by the United Kingdom (22%), Germany (10%), France (8%), Spain (7%) and The Netherlands (6%). Figure 4.2 shows year by year the number of recipients of the financial aid. Between 2007 and 2009, the number of recipients increased significantly as a result of the additional resources from Colciencias. The foundation was able to gather more funds to sponsor more students. However, there is no evidence of an increase in the number of applications in 2007. Figure 4.1 shows the time trends in the number of applicants selected and non-selected as well as those who became recipients of the financial aid. Overall, there is a smooth rise in the total number of applicants between 2004 and 2018.

Figure 4.2: The number of applicants who received the financial aid from Colfuturo from 1992 to 2018



Source: own calculations based on Colfuturo's administrative records.

The beneficiaries receive the amount of money requested during the application process. The money can be used for the payment of the study fee, living expenses, transportation costs from Colombia to the country of destination, books, health insurance, computers, software, among others. Only upon return to Colombia, a percentage of the capital is "forgiven". In the case of receiving a master's/doctorate degree in areas different from management/business, fifty percent (50%) of the capital is forgiven upon return. For the case of receiving an MBA degree or any degree in the area of management/business, a twenty-five percent 25% of the capital is forgiven. Furthermore, an additional 10% is given if the beneficiary works as a public official in an entity of the national, departmental, or municipal order; or full-time teacher or researcher in an educational institution in Colombia. It should be noted that these percentages changed under the CONPES document 3835 from July 2015 and 3862 from July 2016 (DNP, 2015, 2016). For the subsequent cohort of applicants, the percentages of the ”forgiven capital” were substantially reduced. The above influenced the sample selection for this paper.

How relevant is Colfuturo's program when compared to other scholarships available to Colombian graduates? The following table shows the total number of scholarships, credits, and loans-scholarship for master's and doctorates given to Colombian graduates between the period 2004 and 2013.[11] The table shows as well the share by the different institutions. While ICETEX is the largest sponsor of Colombian graduates (44%), Colfuturo represents 24.6 % of the total. Colciencias, through their specific scholarship programs, has issued 25.43% of the total financial aid between 2004 and 2013 (excluding Colfuturo funds). Unlike Colfuturo's scholarship/loan, the credits offered by ICETEX must be paid in full by the student once they have completed their studies. It is also common for students to combine different sponsors to fund more expensive study programs. Nevertheless, Colfuturo represents an important source of funding for many Colombian graduates.

[11] There is no data available for DAAD scholarships. However, the institution is also an important donor of scholarships.

Table 4.1: Scholarships, credits and credits-scholarship for master's and doctorates to Colombian graduates by different institutions 2004 - 2013

Institution	Total Scholarships	Share (%)
ICETEX	9,493	44.19
COLCIENCIAS	5,463	25.43
COLFUTURO	5,285	24.60
Fullbright	519	2.41
MAEC-AECID* Ecopetrol-ICP	188	0.87
Central bank	157	0.73
Fullbright-DNP**Mazda foundation	74	0.34
British government	38	0.17
Total	21,480	100

Source: DNP (2015).*Ministry of Foreign Affairs and Cooperation of Spain - Spanish Agency for International Development Cooperation.** National Planning Department

4.3 Data and Descriptive Statistics

4.3.1 The data

I exploit the administrative records provided by Colfuturo for the cohorts of students who applied to the program in 2011, 2012, 2014, and 2015. I restrict the sample until 2015, given the substantial changes in the scholarship component of the program after 2016, as explained in the previous section.[12] During the selection process, all the applicants must submit a survey via an online platform. The survey includes questions regarding their socioeconomic characteristics, professional profile, information about the university and the program they are planning to study, a budget for the entire duration of the study plan, as well as, all the documentation required for the application. Once the selection process has concluded, the beneficiaries of the program must frequently update their information in the

[12]For 2013, the provided data does not include the identity of the applicants. Given the "Data Use Agreement" between the author and Colfuturo, it is not possible to share the micro-data without their authorization.

online platform to receive the money disbursement (every two months). I use this information for the beneficiaries of the program collected by Colfuturo. To obtain information about the non-beneficiaries of the program, I used information from LinkedIn and CvLAC (Curriculum Vitae of Latin-American and the Caribbean) profiles. The profiles were found after performing thousands of automated queries through Google's Search Console combined with manual validation of the information, comparing the profiles found with the information gathered by Colfuturo during the application process.

The final sample constitutes the applicants to Colfuturo's scholarship program who applied to study for a master's degree. I have deleted all the applicants who planned to pursue a doctoral degree since the number of applicants per study area is small, leading to very high acceptance rates. As noted, I obtained data using other sources for the non-beneficiary applicants. The non-beneficiary applicants include the non-selected applicants and the applicants that resigned to the financial aid. I was able to collect information from online sources for 63% of the total non-beneficiary applicants. Tracking migrants is always challenging in these studies. However, the tracking rate of international students for the present study is comparable to other studies such as Oosterbeek and Webbink (2011) with 51% tracking rate and Parey and Waldinger (2011) with a survey response rate of 25%. Table 4.2 compares the baseline characteristics for both the overall applicants and the final sample used in this analysis. There are no important differences in the composition of the sample used for the analysis when compared with the overall population of applicants, except for the variable university during the bachelor studies. The percentage of applicants who studied in a public university is 1.2 percentage points lower in the sample when compared to the overall population of applicants.

4.3.2 Descriptive statistics

Table 4.2 presents descriptive statistics of all the applicants to the financial aid as well as the sample retained for the main analysis. Columns (1) to (3) show the means and percentages of the baseline characteristics for all the applicants to the scholarship. Columns (4) to (6) show the means and percentages for the primary analysis sample of 5,478 applicants. Table 4.2 shows that nearly all the recipients of the financial aid completed a master degree abroad (98%), while for those who did not receive the money support, approximately 63% finished a master degree abroad up to 4 years after application to the Colfuturo's program [13]. On average, applicants are 27 years old, 49% are female, and a large majority (75% in the sample) received a bachelor's degree from a private university. The reasons behind the under-representation of public university graduates could be that universities abroad select applicants from universities that perform well in global university rankings. Another reason could be that the students with important financial difficulties do not apply to the financial aid.

Moreover, Table 4.2 also shows that 50% of the applicants originate from Bogotá.[14] The city contributes the most to the national GDP (both nominal and per capita), and it is the city with the highest number of universities in Colombia (114 universities). The second region with the most number of applicants is the Andean region, followed by the Caribbean, Pacific, Orinoquia, abroad, and finally the Amazonian region with only 0.3 % of all the applicants.

The table also presents the area of study of the applicants. Most applicants seek to pursue a Master in Management or a Master in Business Administration (MBA). This area

[13] For those who applied multiple times to the call, I selected the last application year; therefore, I consider whether the applicant finished a master up to 4 years after the last application to the program.

[14] For this reason, Bogotá is included as a single category and not unified with the Andean region

of study represents 23% of all the applicants; however, it is also the study area with the highest rejection rates to obtain financial aid. It represents 35% of the non- selected to receive funding, while 16% among the selected. The second-largest study area of interest of the applicants is Engineering (20%), followed by Social Sciences (9.9%), Architecture (8.7%), and Law (7.9%). These five areas of study represent approximately 70% of all the applications to financial aid. It is noted that health is the study area with the highest acceptance rates. Moreover, 9% of all the applicants applied multiple times to the call. A comprehensive discussion about multiple applications is reviewed in the following section.

4.4 Identification Strategy

This paper aims to study the effect of financial aid on studying abroad by analyzing a sample of applicants participating in Colfuturo's Scholarship-Loan program. A simple comparison of students who receive an international study aid versus those who did not obtain it could lead to a biased estimation of this effect since the two groups of students could be different in both observed and unobserved characteristics. Students who typically apply and obtain international study grants are high-achieving students, interested in living abroad, more entrepreneurial, risk-takers, and interested in other cultures. Most of these determinants are difficult to account for. For example, let's say that those who are the most motivated are the ones who apply and obtain the grant are compared to students who did not obtain it. The latter group includes those who did not apply as well as a portion of those who did apply but did not receive the financial aid. Those who did not apply might not be motivated into studying abroad, decreasing the likelihood of studying abroad for the overall group of non-treated individuals. For this reason, the identification strategy is based on the discontinuity of the treatment occurring at the cutoff point defined in the selection process. The cutoff

Table 4.2: Descriptive statistics - Colfuturo sample

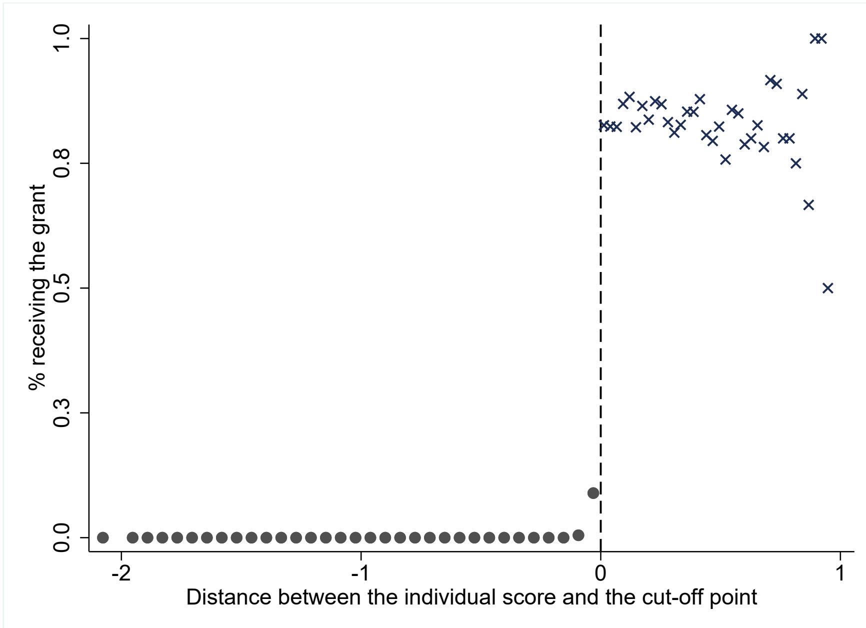
	All applicants			Sample		
	Not-Selected	Selected	Total	Not-Selected	Selected	Total
	(1)	(2)	(3)	(4)	(5)	(6)
Master Abroad				63.4	98.5	86.3
Women	47.5	49.9	48.9	47.1	50.1	49.2
Age	27.5	27.1	27.2	27.3	27.1	27.2
Undergraduate university						
-Private	76.8	72.0	73.9	79.5	73.2	75.1
-Public	23.2	28.0	26.1	20.5	26.8	24.9
Region born						
-Abroad	1.4	1.7	1.6	1.4	1.7	1.6
-Amazonia	0.2	0.3	0.3	0.2	0.3	0.3
-Andean	30.8	29.8	30.2	29.5	28.9	29.5
-Bogotá	50.9	52.1	51.6	51.0	52.3	51.9
-Caribean	7.8	7.5	7.6	8.5	7.7	8.0
-Orinoquia	1.1	0.8	0.9	1.2	0.8	0.9
-Pacific	7.8	7.8	7.8	8.3	7.8	7.9
Area of Study						
-Agriculture&Environment	1.4	6.0	4.2	1.5	5.9	4.3
-Architecture	8.9	8.6	8.7	8.8	8.7	8.8
-Arts	5.7	7.6	6.9	4.5	7.7	6.7
-Basic	1.1	5.1	3.5	0.9	4.6	3.4
-Economics	2.5	2.9	2.7	2.4	3.0	2.8
-Education	0.8	3.7	2.6	0.5	3.6	2.6
-Engineering	20.1	20.4	20.3	19.7	20.2	20.0
-Health	1.5	4.1	3.0	0.9	4.0	3.0
-Law	6.9	8.6	7.9	7.0	9.0	8.4
-Management	35.0	16.0	23.6	38.9	16.2	23.4
-Political	5.4	7.7	6.8	5.6	8.0	7.2
-Social	10.8	9.3	9.9	9.2	9.0	9.1
Multiple applications	10.4	8.0	8.9	12.0	7.9	9.2
Observations	2,734	4,109	6,843	1,723	3,755	5,478

Notes: The table show percentages for most variables except age, for which the table shows the average. Multiple applications is a dummy that takes the value of 1 if the applicant submitted an application in previous years and zero otherwise point separates the two groups exogenously if this point has not been manipulated by any party. In principle, these differences between the groups are negligible when I compare selected and not selected applicants with scores close enough to the cutoff point.

The design of Colfuturo's program allows eliminating potential selection bias since it enables the comparison of applicants with equal interest in studying abroad, similar academic abilities, as well as other observable and unobservable characteristics. By restricting the sample to those around a very narrow bandwidth of both sides of the cutoff point, I can identify the treatment effect of financial aid on the probability of studying abroad. The regression discontinuity design (RDD) analysis has been used widely used in the education literature (e.g. Imbens and Lemieux, 2008; Lee and Lemieux, 2010) and in the international migration literature (Pinotti, 2017; Lochmann et al., 2019). For example, Thistlethwaite and Campbell (1960) studied the effect of merit awards on subsequent academic outcomes using RDD.

While the treatment is mostly available for those who score above the cutting line, not all selected applicants are treated. Figure 4.3 presents the percentage of applicants who receive the financial aid in relation to the normalized score. A fraction consisting of 16% of the students were selected into treatment but did not accept the financial aid. Colfuturo records different reasons that explain why the selected applicants rejected the financial aid. Reasons such as health, family, visa denial, and other personal reasons can be observed in Coltufuro's records. However, in the majority of the cases, the reason for resignation is unknown. There are only ten recorded cases for which the applicant was selected but did not fulfill the co-debtors requirement to received the Scholarship/loan. Moreover, 0.48% of the applicants in the sample received the treatment despite not being selected by the second committee. One potential reason could be the availability of additional funding. Overall,

Figure 4.3: Receiving the financial aid from Colfuturo and the score



Source: own calculations based on Colfuturo's administrative records

eligible individuals self-selected into the treatment since the participation in the program is voluntary, and some non-eligible individuals selected themselves into the treatment. In this case, the probability of receiving the financial aid increments discontinuously at the cutoff point (as depicted in figure 4.3). Considering the fuzzy nature in this scenario, the following two equations were estimated under the instrumental variable approach:

$$B_i = \alpha_0 + \alpha_1 C_i + F_1(S_i - c) + X_i\gamma + \epsilon_i \tag{4.1}$$

$$Y_i = \beta_0 + \beta_1 B_i + F_2(S_i - c) + X_i\theta + \mu_i \tag{4.2}$$

The first equation models the probability of receiving financial aid. Let B_i=1 for the applicant who is a beneficiary of the Colfuturo Scholarship-Loan (received the financial aid) and B_i=0 for those students who did not receive the Colfuturo aid. The variable S_i denotes the score obtained by the students during the selection process. S_i is also the assignment or "running" variable through which the selection committees define who is eligible or not for receiving the financial aid. This variable is centered at point c, which is the cutoff point. Let C_i =1 for students who were selected to receive the Colfuturo's Scholarship-Loan, and C_i=0 those students who were not selected to receive the financial aid. If $C_i \equiv 1[S_i \geq c]$ denotes, the individual i was selected into treatment since the score is higher than the cutoff point. If $C_i \equiv 0[S_i < c]$ denotes the individual i was not selected into treatment. Being selected for treatment increases the probability of becoming a beneficiary of the program.

The second equation estimates the probability of studying abroad, where Y_i represents whether the applicant finished a master's degree abroad between one to four years after the application was sent to Colfuturo.[15] Studying abroad then is a function of receiving the financial aid (B_i), the normalized score ($S_i - c$), X_i which corresponds to the vector of explanatory variables included to improve the precision of the estimates and $F(S_i - c)$ which is a polynomial function.

4.4.1 Validity checks

As pointed out in the previous section, this paper analyses the context of Colfuturo's scholarship/credit program under the RDD approach. One potential issue that could jeopardize the internal validity of the study is the suspect of manipulation. In this context, an applicant's rank depends on how she or he compares to other applicants. It is nearly impossible for an applicant to have information about all the other applicants before applying to the

[15]The estimated effects are conditional on attendance. I can observe from the sample of recipients on 41 cases that couldn't complete their studies; however, there is no information about the non-beneficiaries who went abroad but did not finish the studies.

program. Moreover, the final selection depends on the quality of the other applications in the same year and area of study; therefore, the candidate cannot know in advance his or her relative ranking compared to the other applicants. In addition, the cutoff rank belonging to to a specific year and study area is unknown in anticipation, and it changed every year. The cutoff depends on the number of applicants in the study area and the number of scholarships available for that year.

It is important to note that students can participate as many times as preferred to the application call. The above is not a threat to the identification of the estimates because the total score in the selection process depends on the four criteria that can hardly be manipulated by the applicant. The first two criteria (average grade point average (GPA) obtained in the bachelor's studies and the relative ranking within their graduating cohort in the bachelor's studies) cannot be altered by the student after finishing the studies. [16] The student can improve his or her writing skills and apply to a well-ranked university in the next year, however, the score obtained for these two criteria will depend on being selected in a well-ranked university the next year and the reviewers of the essay. From the 18,903 applicants between 2010 and 2018 applicants, approx. 9% of those sent their application more than once. For these applicants, I focus on the last attempt to participate in the program between 2010 and 2018. In our final sample, approximately 57% of the applicants who applied to the call more than once were selected for the treatment, compared to 43% who were not selected for treatment.

[16]Each applicant must provide a certified copy of the bachelor's degree, the university transcripts, and a letter describing the relative rank within their graduating cohort.

Table 4.3: Test of Balance of Baseline Characteristics

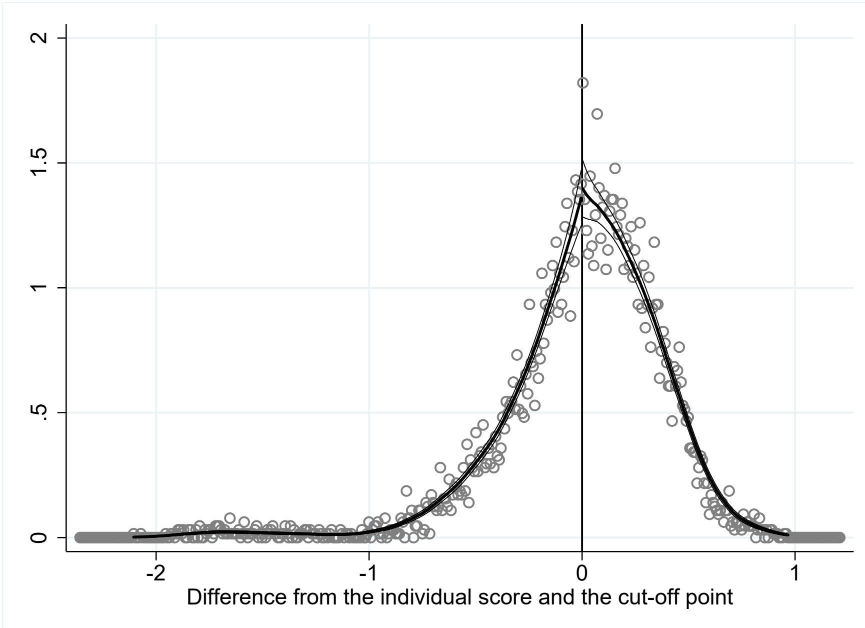
	(1) Age	(2) Male	(3) Public University	(4) Multiple applications	(5) Origin: abroad	
I(Score $\geq$ Cutoff)	0.0875	0.0101	0.0564*	-0.0348	-0.00185	
	(0.278)	(0.0284)	(0.0192)	(0.0315)	(0.00411)	
Polynomial degree	2	2	2	2	2	
Observations	5475	5475	5475	5475	5475	
	(6)	(7)	(8)	(9)	(10)	(11)
Region of origin	Amazonia	Andean	Bogota	Caribean	Orinoquia	Pacific
I(Score $\geq$ Cutoff) selected	-0.000662	-0.000852	0.0273	-0.0217*	0.00383	-0.00530
	(0.00243)	(0.0212)	(0.0270)	(0.00971)	(0.00388)	(0.0129)
Polynomial degree	2	2	2	2	2	2
Observations	5475	5475	5475	5475	5475	5475
	(6)	(7)	(8)	(9)	(10)	(11)
Subject of study	Arts	Enviroment	Architecture	Basic	Economics	Education
I(Score $\geq$ Cutoff)	0.0165	0.00405	-0.00117	0.0169	0.00936	0.0262
	(0.0159)	(0.00646)	(0.00849)	(0.0148)	(0.0100)	(0.0227)
Polynomial degree	2	2	2	2	2	2
Observations	5475	5475	5475	5475	5475	5475
	(6)	(7)	(8)	(9)	(10)	(11)
	Engineering	Health	Management	Law	Political	Social
I(Score $\geq$ Cutoff)	-0.0168	0.0305	-0.0715	-0.0152	0.000524	0.000689
	(0.0284)	(0.0321)	(0.0470)	(0.0159)	(0.00232)	(0.00521)
Polynomial degree	2	2	2	2	2	2
Observations	5475	5475	5475	5475	5475	5475

Notes: Robust standard errors in parentheses, clustered by subject of study. Each regression includes functions in the overall score of polynomial degree 2 and interactions with I(Score $\geq$ Cutoff), which is a dummy variable that takes the value of 1 if the score exceeds the cutoff and zero otherwise. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To test whether there was manipulation in the assignment of the financial aid, I analyze the students' baseline characteristics to check for any discontinuities in the observed covariates around the cutoff. First, the score or "running" variable was divided into 36 bins, and a simple mean for each baseline variable is calculated for the observations within each bin. Figures 4.8 and 4.9 graphically present the means for each variable at different points of the score. The graphic analysis does not show any discontinuity around the cutoff for the observed characteristics of the applicants. Second, I estimate different unrelated regressions in which the dependent variables are each baseline characteristics regressed on the dummy variable that indicates having a score higher to the cutoff point (C_i), the score (S_i), a quadratic polynomial of S_i and interactions between C_i, the score and the quadratic polynomial. The results summarized in Table 4.3 show that most covariates do not change discretely at the threshold, except for the one coefficient associated with the type of university during the bachelor studies, which is statistically significant. However, the coefficient is substantively small (0.05); therefore, I can not conclude there is systematic sorting around the cutoff based on one variable.

To test further for potential manipulation, I follow the proposed test by McCrary (2008). In a nutshell, the author formulates a test to examine whether the density function of the running variable jumps in the vicinity of the cutoff point. Given that the selection process involves many different cutoffs (by area of study and year), I normalized the score and the cutoffs to zero. The density function of the normalized score or "running" variable is displayed in Figure 4.4. The results presented in Figure 4.4 do not show evidence of bunching of the running variable (or the normalized score assigned by the selection committees) around the cutoff point.

Figure 4.4: The density of the normalized score

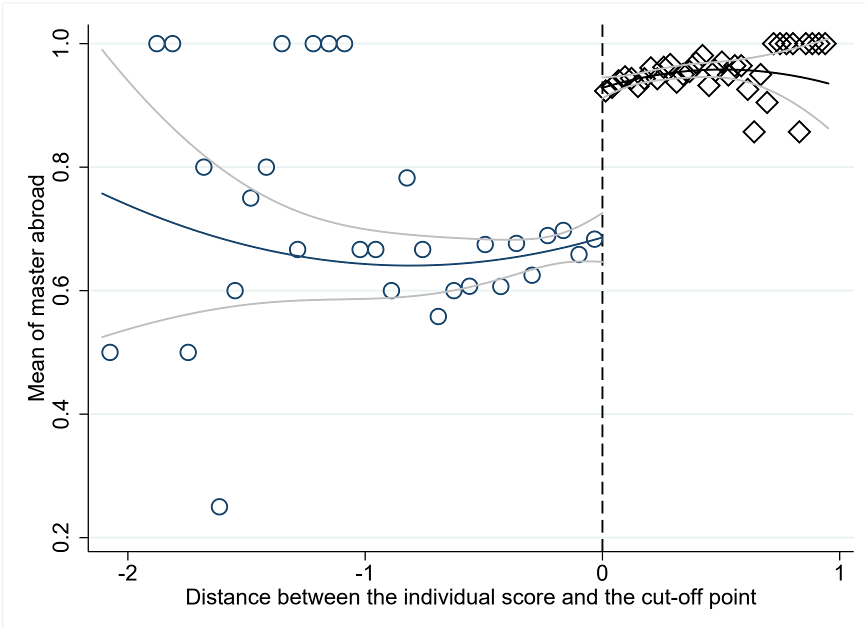


4.5 The empirical findings

4.5.1 Baseline estimates

In order to examine the relationship between receiving Colfuturo's financial aid and the changes of obtaining a master's degree abroad, I first inspected the raw data graphically. The same procedure described in the previous sections is followed for the outcome of interest. The score was divided over 35 bins of similar width, and the percentage of applicants who finished a master's degree abroad is calculated for the observations within each bin. Figure 4.5 displays the calculated percentages at each bin along with the normalized score. As expected, applicants with the highest scores were very likely to study abroad since they were students with excellent grades. Moreover, figure 4.5 shows a clear discontinuity around the cutoff point in the percentage of applicants who completed a master's degree abroad. On

Figure 4.5: The percentage of applicants who received a master's degree abroad



the right side, just above the cutoff, the percentage of applicants who completed studies abroad is around 92%. On the left side, just below the cutoff, the percentage of applicants who receive a master's degree overseas is nearly 68%. The figure also displays the best fit for the data, which is a quadratic function. As part of the sensitivity analysis, I also test for different polynomials (see figure 4.7).

To test statistically whether Colfuturo's financial aid increases the chances of successfully completing a master's degree abroad, equations 4.1 and 4.2 are estimated both using non-parametric and parametric methods. To estimate the effect using two-stage local polynomial regression, it is required to choose the sub-sample of applicants just above and below the cutoff. To select the neighborhood observations around the threshold, I utilize a data-driven optimal bandwidth selector developed by Calonico et al. (2014). The method developed by Calonico et al. (2014) is also adaptable to the fuzzy RDD.

Table 4.4: Local linear estimates of the effect of the scholarship/loan on the probability of obtaining a master degree

	(1)	(2)	(3)
Panel A: equation 4.2		Master degree abroad	
Colfuturo beneficiary	0.324***	0.309***	0.311***
	(0.0626)	(0.0593)	(0.0434)
Panel B: equation 4.1		Colfuturo beneficiary	
I(Score $\geq$ Cutoff)	0.59461***	0.61217***	0.60908***
	(0.02689)	(0.02582)	(0.06115)
Polynomial of degree	1	1	1
Control variables	No	Yes	Yes
Study subject Fixed Effects	No	Yes	Yes
Year Fixed Effects	No	Yes	Yes
Clustered SE	No	No	Yes
Eff. obs. left of c	528	543	531
Eff. obs right of c	1018	1042	1022
Observations	5475	5472	5472

Notes: Robust standard errors in parentheses for columns (1) and (2). For column (3) the standard errors were clustered by are of study. Control variables include gender, age, age squared,a dummy variable that indicates if the student applied more than once, country of destination, application year, region of origin within Colombia and a dummy variable that takes the value of 1 if the applicant studied the bachelor's degree in a public university and the study area. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Panel A of the table 4.4 presents the estimation results for equation 4.2 using a polynomial of degree 2 over the sample selected using local polynomial regressions. Column (1) presents the estimations without including any control variables. The results shows that receiving Colfuturo's funding increases the probability of completing a master's degree abroad by 32 percentage points. Column (2) includes the following control variables: gender, age, age squared, a dummy variable that indicates if the student applied more than once, country of studies[17], year of application, region of origin within Colombia or abroad, and a dummy

[17]For those who did not study abroad, I use the country where they planned to study.

variable that takes the value of 1 if the applicant studied a bachelor's degree in a public university. After the inclusion of these control variables, the coefficient changes slightly to 0.309.

Moreover, column (3) of the table 4.4 presents the estimations when the standard errors have been clustered by area of study to account for any potential correlation among individuals that applied within the same study area. Overall, the results confirm a statistically significant impact of receiving Colfuturo's scholarship/loan on the probability of studying abroad for Colombian graduates. Panel B presents the results for the equation 4.1 over the sample in the neighborhood of the threshold. Unsurprisingly, the results confirm that being selected to receive the aid or namely having a score above the threshold increases the probability of being a beneficiary of the financial aid offered by Colfuturo.

To estimate the effect of Colfuturo's scholarship/loan using parametric estimations methods, I make use of the entire sample and add polynomials up to the second degree (quadratic) following Gelman and Imbens (2019). Gelman and Imbens (2019) argue that parametric estimations can be sensitive to the degree of the polynomials. The authors demonstrate that higher degree polynomials can lead to noisy weights and misleading confidence intervals. Table 4.5 summarizes the estimation results of equations 4.1 and 4.2 using Two-Stage Least Squares (2-SLS). [18] Once again, the table presents in column (1) the results without control variables, column (2) includes the same control variables, and column(3) includes all the control variables and clusters the standard errors by area of study. Across specifications, results using parametric estimations confirm that being a beneficiary of the financial aid offered by Colfuturo increases the probability of studying abroad. The 2-SLS results are

[18]While our dependent variable is certainly a dichotomous variable, the endogenous variable is also dichotomous, therefore, estimating both equation 4.1 and 4.2 using a Two-Stage procedure is considered in the literature as the "forbidden regression".

Table 4.5: Parametric estimates of the effect of the scholarship/loan on the probability of obtaining a master degree

	(1)	(2)	(3)
Panel A: Second stage	Master degree abroad		
Colfuturo beneficiary	0.285***	0.292***	0.292***
	(0.0213)	(0.0213)	(0.0341)
R^2	0.233	0.254	0.254
Panel B: First stage	Colfuturo beneficiary		
I(Score $\geq$ Cutoff)	0.938***	0.935***	0.935***
	(0.0133)	(0.0135)	(0.0094)
R^2	0.872	0.875	0.875
First stage F-stat.	9882.88	9738.11	
Polynomial of degree	2	2	2
Control variables	No	Yes	Yes
Study subject Fixed Effects	No	Yes	Yes
Year Fixed Effects	No	Yes	Yes
Clustered SE	No	No	Yes
Observations	5475	5472	5472

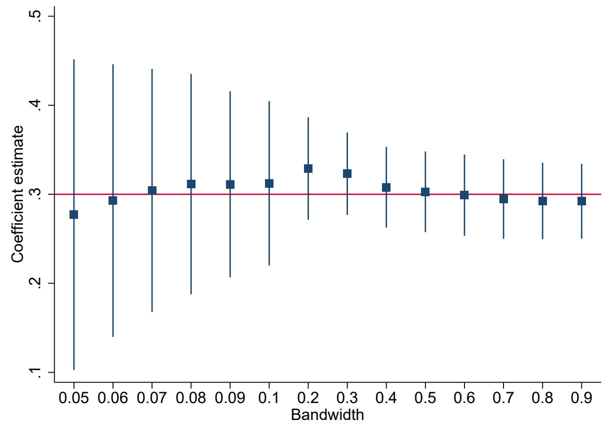
Notes: Robust standard errors in parentheses for columns (1) and (2). For column (3) the standard errors were clustered by subject of study. Each regression includes functions in the overall score of polynomial degree 2 and interactions with I(Score $\geq$ Cutoff), which is a dummy variable that takes the value of 1 if the score exceeds the cutoff and zero otherwise. Control variables include gender, age, age squared,a dummy variable that indicates if the student applied more than once, country of destination, application year, region of origin within Colombia and a dummy variable that takes the value of 1 if the applicant studied the bachelor's degree in a public university and the area of study. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

independent of the inclusion of other covariates since the coefficients change slightly from 0.28 to 0.29. Overall, it is plausible to conclude that Colfuturo's scholarship/loan is an effective tool to increment the probability of studying abroad for Colombian graduates by 29 to 30 percentage points. Panel B presents the results for the first stage of 2-SLS. The results using the overall sample show a higher coefficient when compared with Panel B of Table 4.4. Overall, it is not surprising that being selected into the treatment is a good predictor for being a recipient of the financial aid offered by Colfuturo.

4.5.2 Sensitivity checks

Under the local polynomial regression approach, the selection of the bandwidth is crucial. Wider bandwidths allow for a larger number of observations to be included in the estimations and more precise estimates as a result. However, a wider bandwidth could lead to bias estimates since the local randomization assumption might not hold. To test the sensitivity of the results to the selection of the bandwidth, I re-estimate equations 4.1 and 4.2 under different bandwidths. The selected bandwidths are symmetrical, meaning the same width is used for the data below the cutoff and the data above the cutoff. Figure 4.6 presents the coefficient of receiving Colfuturo's aid in equation 4.2 including all control variables, interaction with the S_i and C_i, and the clustering of the standard errors by area of study for different sizes of the neighborhood around the threshold. The results presented in table 4.6 exemplify the trade-off between precision and bias under the local polynomial regression approach. When the bandwidth is small, the confidence intervals are larger, and the point estimate is less precise. As the bandwidth becomes larger, the confidence intervals are narrower, and the coefficient is estimated with more precision.As the bandwidth increases, the results become similar to the estimates presented in table 4.5 using parametric estimations. Nevertheless,

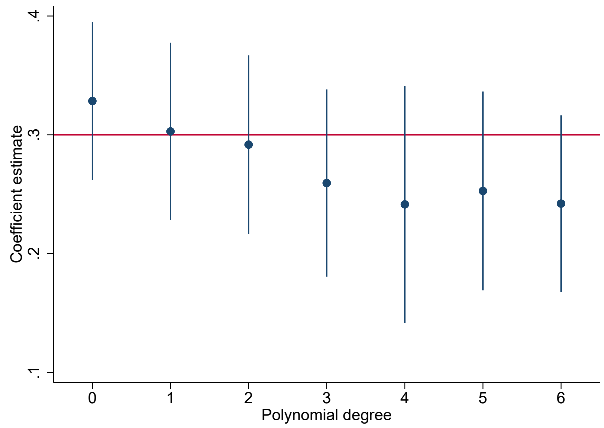
Figure 4.6: Non-parametric estimates of the probability of completing a master degree abroad using different bandwidths



the results presented in figure 4.6 confirm the relevance of Colfuturo's aid on completing studies abroad.

A different approach is presented to assess the sensitivity of the estimations under the parametric methods. In this case, the interest focuses on finding the best model that fits the given data. For this reason, I estimated equations 4.1 and 4.2 using higher-order global polynomials of degrees between 1 and 6, as well as the interaction with S_i and C_i, the inclusion of control variables, and standard errors clustered at the study area. The results displayed in Figure 4.7 confirm the discussion presented by Gelman and Imbens (2019) about the use of high order polynomials for the running variable. The results are sensitive to controlling for high-degree polynomials for the continuous "running" variable. Notwithstanding, the estimated coefficient oscillates between 0.24 and 0.302. Overall using both different polynomial degrees and bandwidths, the estimated coefficient of interest is stable, and it fluctuates around 0.3.

Figure 4.7: Parametric estimates (2SLS) of the probability of completing a master degree abroad using alternative polynomial function



4.6 Discussion and Conclusions

Numerous developing countries have expanded the number of government-sponsored scholarships to fund studies abroad for their citizens. For example, CAPE scholarship for Brazilian graduates, DIKTI scholarship in Indonesia, Bolashak Scholarship in Kazakhstan, CONACYT scholarship in Mexico, Scholarship No. 322 in Vietnam, among many others. The financial support of students in studying abroad has become highly relevant for countries as a mechanism to achieve high-quality tertiary education among their population. The case of the Colfuturo Scholarship - Loan program is not an exception.

The Colfuturo Scholarship - Loan program aims at strengthening the formation of highly qualified human capital at the best universities in the world. The program funds master and Ph.D. degrees among Colombian graduates through a specific financial aid that mixes different elements. It consists of a conditional scholarship that depends on a return criterion. If the students decide not to return to Colombia, the money disbursed during the study times transforms into a study loan. The program has been very successful by sponsoring over 17 thousand students between 1992 and 2020. Moreover, the structure of the program generates incentives to return to Colombia. According to Colfuturo's records, on average, 74% of the recipients return to Colombia after the study period.

This paper aims to estimate, causally, the role of financial aid such as Colfuturo's program on the probability of studying abroad. Given the characteristics of the selection process, I implemented a Regression Discontinuity Design (RDD). The present paper contributes to the literature related to the determinants of studying abroad by accounting for selection bias as a result of financial aid self-selection. To summarize, the results suggest that being a recipient of Colfuturo's financial help increased the probability of studying abroad at the post-graduate level by 30 percentage points approximately. The results are very similar to those found by Oosterbeek and Webbink (2011). The authors show that receiving the grant "program for the talented" increases the probability of studying abroad by 24.3 to 29.5 percentage points depending on the specification.

The estimated effects of financial aid on studying abroad are larger in magnitude when compared to the effect of other types of financial aid to study within the country of origin. For example, Dynarski (2003b) calculates that the effect of the eligibility to the Social Security Student Benefit Program increases on college attendance by 21.9 percentage points and completion by 14.5 percentage points. Dynarski (2008) estimates the effect of two programs: the Georgia and Arkansas merit scholarship on degree completion, and found an effect equal to 3-4 percentage points. Moreover, using the PROMISE program Scott-Clayton (2011) found that receiving financial aid increases four-year-bachelor completion by 9.4 percentage points.

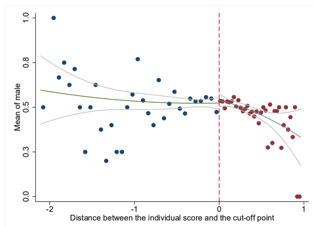
The estimated effects of Colfuturo's financial aid on studying abroad in thousands of dollars are lower in magnitude when compared to the results of other studies. The estimated effect can be easily translated into costs. Colfuturo paid to its recipients on average 33,606 US dollars. This means that for every 1,000 US dollars spent per student, the probability of successful completion of a master's degree abroad increases by 0.89 percentage points. In comparison, Whatley (2017) found that every 1,000 US dollars received in grant aid increase the probability of studying abroad by 29 percentage points among the students in the university system of the state of Georgia, USA. Moreover, Dynarski (2000) and Cornwell et al. (2006) found that for every 1,000 US dollars given by Georgia's HOPE scholarship, college enrollment of freshmen students increases by 4 to 6 percentage points. Nevertheless, one must be cautious in generalizing the results obtained for Colfuturo's program and compare it to any financial aid. Given the characteristics of the Colfuturo Scholarship/loan, it is not simple to disentangle the effects of the loan part from the scholarship part. Another limitation of this study lies in the fact that the estimated effects are conditional on attendance to a university abroad since it is not possible to obtain information of who among the non-beneficiaries that did not obtain a master's degree abroad tried but did not complete the degree. Moreover, there is potential sample selection since the participants in the selection process could not be representative of the Colombian population. Higher education is not widely accessible in Colombia. According to OECD (2020a), 30% of the population in Colombia obtained a tertiary degree. This percentage is low when compared to the OECD average (45%).

Future research could explore the potential heterogeneous effects of the program to understand who are the most benefited by the financial aid. Are students coming from public universities most benefited from the program? Or does financial aid impact a student's decision to attend a year versus two-year master program or public versus private university abroad?

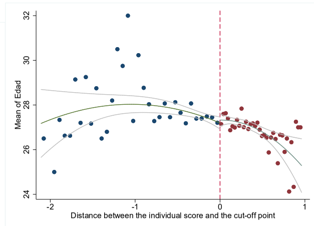
4.7 Appendix C

4.7.1 Appendix C1. Graphical analysis of the baseline characteristic around the cut-off point

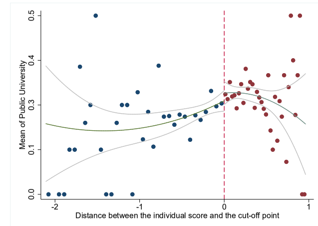
Figure 4.8: Graphical analysis of the baseline characteristic Colfuturo's recipients - Part 1.



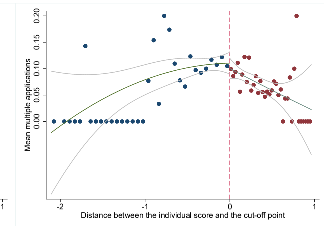
(a) Male



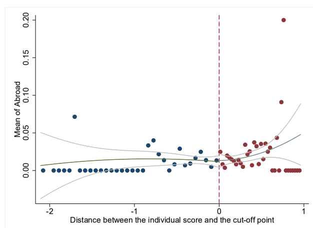
(b) Age



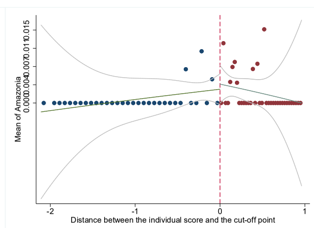
(c) Public University



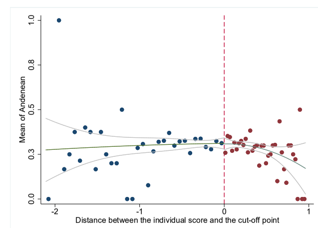
(d) Multiple applications



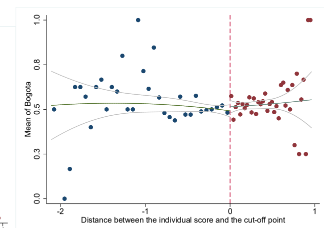
(e) Origin: Abroad



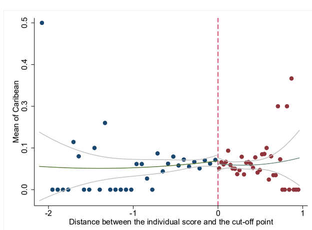
(f) Origin: Amazonia



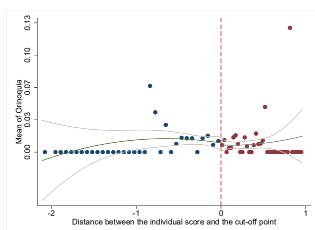
(g) Origin: Andean



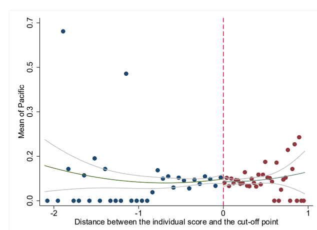
(h) Origin: Bogota



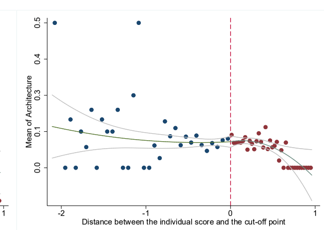
(i) Origin: Caribean



(j) Origin: Orinoquia

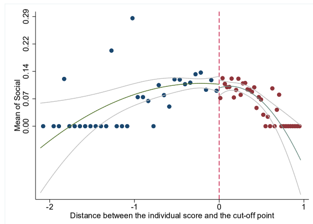


(k) Origin: Pacific

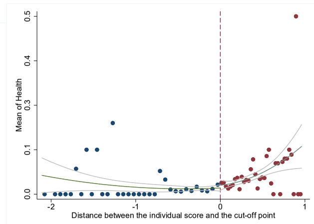


(l) Area: Architecture

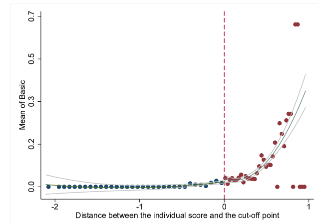
Figure 4.9: Graphical analysis of the baseline characteristic Colfuturo's recipients - Part 2.



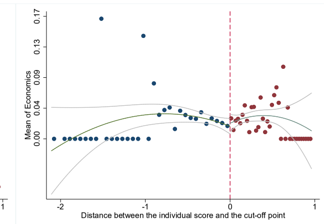
(a) Area: Social studies



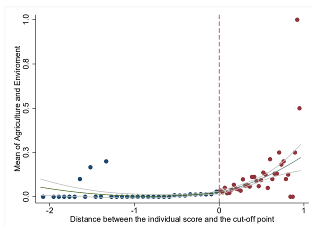
(b) Subject: Health



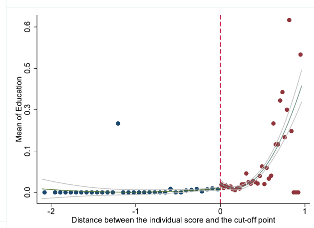
(c) Area: Basic Science



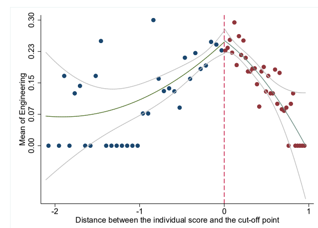
(d) Area: Economics



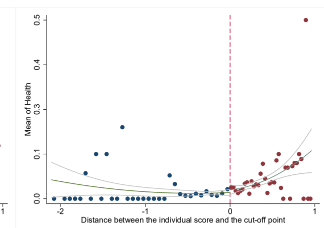
(e) Area: Agriculture and Enviroment



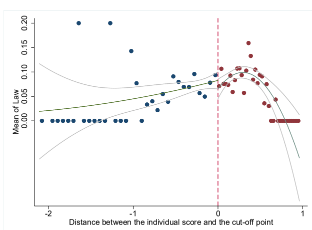
(f) Subject: Education



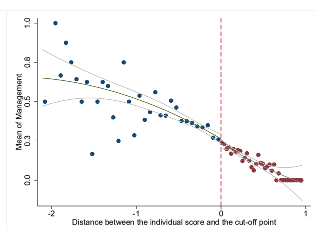
(g) Area: Engineering



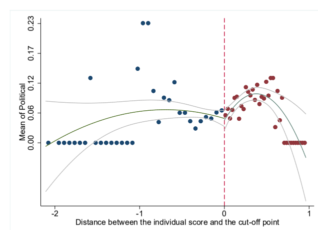
(h) Area: Health



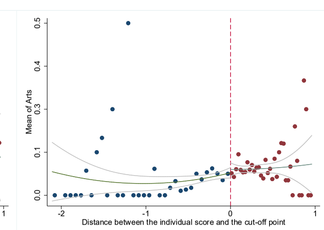
(i) Area: Law



(j) Area: Management



(k) Area: Political



(l) Area: Arts

CHAPTER 5
CONCLUSION

The purpose of this thesis is to give light to specific questions regarding international migration. It covers aspects about the determinants that explain people's movements abroad and the integration process in the host societies. Each chapter of this dissertation fills different research gaps in the literature. The next section session summarizes them.

5.1 Summary of key findings and significance

5.1.1 Migration Diasporas and Long-term Cultural Distance: Testing the Collier's hypothesis

The effect of the stock migrants in a country on the subsequent inflows of migrants has been widely documented in the literature. This effect is known as the network effect. The purpose of chapter 2 is to understand whether the network effect on migration flows is larger when the cultural distance between the country of origin and destination is large. This objective is accomplished by estimating the interaction between the network effect and ancestral distance on international migration flows. For this task, a gravity model to explain international migration inflows was estimated using the Poisson Pseudo Maximum Likelihood (PPML) estimator developed by Santos Silva & Tenreyro (2006). The estimated elasticity of migration stocks on migration flows ranges from 0.6 to 0.8 depending on the level of ancestral distance. This difference of 0.2 percentage points in the inflow of migrants between culturally distant and culturally similar country pairs is rather small considering the length of time of the estimates (10 years).

5.1.2 The solution of the immigrant paradox: aspirations and expectations of children of migrants

In contrast to chapter 2, chapter 3 studies the educational performance of the children of migrants in the U.S. It shows that the difference in school performance between migrant children and natives lies within the misalignment in aspirations and expectations that migrant children form. The chapter shows that a positive difference between aspirations and expectations is a driving force for higher effort and better education outcomes of immigrant teenagers. This force resolves the well-known immigrant paradox. The result is specific to migrant children and does not hold for second-generation migrant pupils who appear quite acculturated to the USA context. This chapter evidences that when considering a broad set of explanatory determinants of educational attainment, we can fully explain the differences in performance between the children of migrants and natives.

5.1.3 The role of financial aid on foreign education: Evidence from Colombian graduates

This chapter studies the causal effect of receiving financial aid on foreign education at the postgraduate level. I use a sample of Colombian graduates who applied to Colfuturo's financial aid program. The characteristics of Colfuturo's selection process allow the implementation of a Regression Discontinuity Design (RDD). The assignment of the fund is expected to be distributed quasi-randomly for the applicants around the vicinity of a cut-off point. This methodology allows for the estimation of the local average treatment effect (LATE) of the program. The chapter contributes to the literature related to the determinants of studying abroad by accounting for selection bias as a result of financial aid self-selection. To summarize, the results suggest that being a recipient of Colfuturo's financial help increased the probability of completing a master's degree abroad by 30 percentage points approximately. The scholarship-loan program is an effective tool to promote high-quality education among Colombian citizens. The results are remarkably robust across estimation methods.

5.2 Limitations

The chapters in this dissertation are not free of limitations. Each chapter employed different data and statistical methods. The limitations will be covered chapter by chapter.

Unlike previous studies, chapter two builds upon migration flow data from a new source. Traditionally, migration is measured with some accuracy through population census and surveys. These data sources are not only expensive to carry out in some countries but also infrequent and available with delay. These sources of information could estimate migration with some error since they often leave minorities out of the sample or represented by very few observations. To estimate the inflow of international migrants, the literature traditionally employs the difference in the stock of immigrants from one census period to the other. However, this method could measure immigration inflows with a degree of inaccuracy since it does not take into account mortality rates, regularization programs, return migration, or the heterogeneity of agents (Beine et al., 2010a). The second chapter makes use of migrations inflows obtained from population registries, border statistics, and permits (residence/work) instead of changes in the stock of migrants. These sources are not free of measurement error as well. For example, irregular migrants are not accounted when using this type of data source, therefore, the estimates of migration inflows could be seen as a lower bound of the total inflows under high numbers of irregular migrants. Nevertheless, results in chapter

2 show that using these alternative information sources leads to similar estimates of the network effect on international migration flows than when using the differences in migration stocks.

Another limitation is the use of ancestral distance as a measure of cultural distance. This proxy can only capture cultural aspects that are particularly invariant through time. While this is an advantage of the study, it can also be a limitation since it leaves out many cultural aspects that are in continuous transformation. The lack of historical data on cultural values makes this task hard to accomplish. Moreover, the reverse causality between cultural distance and international migration flows presents a challenge if one wants to take into consideration time-variant aspects of cultural distance.

In chapter 3, the use of GPA as a measure of educational performance rather than a standardized test is a limitation of the study. Teachers have complete discretionary power to assign grades to each student. This judgment differs from teacher to teacher and between academic institutions. The above creates a comparability issue across classes and schools. The inclusion of fixed effects partially accounts for this issue, however, the use of GPA is still a limitation in the study. Another important limitation is the sample size of native teens compared to migrant teens. While Add Health contains around 20 thousand observations, the migration sample is still around 20% of the overall sample. The above leads to a large sample for native teens and a small sample for migrant teens. While larger samples increase statistical power, it can also lead to highly significant estimates, especially for native teens in this case. In addition, it should be highlighted that non-enrollment rates in high school are larger among migrant teens than compared to native teens if the migrants arrived late in their teens. The above means that there is scope for selection bias in the sample if we consider those migrant teens who perform well at school are enrolling in high school while those with lower performance might not be observed in the sample.

Lastly, it should be noted that chapter 4 is constricted to study the successful completion of the degree since it was not possible to obtain information of those non-beneficiaries who did not obtain a master's degree abroad but tried to do so. The above means that the estimated effects are conditional on attendance to a university abroad. The analysis is also not considering whether the students obtained other financial aid different from Colfuturo. Given the characteristics of the Colfuturo Scholarship/loan, it is not simple to disentangle the effects of the loan part from the scholarship part. Moreover, there is potential sample selection since the participants in the selection process could not be representative of the Colombian population. Higher education in Colombia is still not widely accessible.

5.3 Opportunities for future research

The exploration in chapter 3 could be extended to other major destinations countries in Europe, where the education systems differ greatly from the education system in the USA. The idea of the *American Dream* as the freedom and ability to write your own future could be also affecting the self-efficacy of migrant children. This view is unique to American society, therefore, the exploration of aspirations, expectations, and misalignment in other contexts is relevant to fully understand the gap in the performance of migrant children and native children.

Furthermore, the study of international students and their migratory trajectories is still an emerging field. Future research using the data from chapter 4 could explore the potential heterogeneous effects of the program to understand who are the most benefited by the financial aid. Are students coming from public universities most benefited from the program? or does financial aid impact a student's decision to attend a year versus two-year master program, or public versus private university abroad? Moreover, understanding whether higher education abroad payoff to the students remains understudied.

BIBLIOGRAPHY

Abramitzky, R., Boustan, L. P., Jacome, E., and Pérez, S. (2019). Intergenerational mobility of immigrants in the us over two centuries. *NBER Working Paper*, (w26408).

Adserà, A. and Pytliková, M. (2015). The role of language in shaping international migration. *The Economic Journal*, 125(586):F49–F81.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8:155–94. O. Galor (ed.) (2011), Inequality and Economic Development: The Modern Perspective, Edward Elgar, UK.

Algan, Y., Dustmann, C., Glitz, A., and Manning, A. (2010). The economic situation of first and second-generation immigrants in france, germany and the united kingdom. *The Economic Journal*, 120(542).

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1):151–184.

Artuc, E., Docquier, F., Çaglar Özden, and Parsons, C. (2015). A global assessment of human capital mobility: The role of non-oecd destinations. *World Development*, 65:6 – 26. Migration and Development.

Bandura, A. (2001). Self-regulation of motivation through anticipatory and self-reactive mechanisms. In *Current theory and research in motivation*, volume 38 of *Nebraska Sympo-*

sium on Motivation 1990: Perspectives on motivation, page 69–164. University of Nebraska Press.

Batista, C., Lacuesta, A., and Vicente, P. C. (2012). Testing the 'brain gain' hypothesis: Micro evidence from Cape Verde. *Journal of Development Economics*, 97(1):32–45.

Beine, M. (2016). The role of networks for migration flows: an update. *International Journal of Manpower*, 37(7):1154–1171.

Beine, M., Bertinelli, L., Cömertpay, R., Litina, A., and Maystadt, J.-F. (2021). A gravity analysis of refugee mobility using mobile phone data. *Journal of Development Economics*, 150:102618.

Beine, M., Delogu, M., and Ragot, L. (2018). The role of fees in foreign education: evidence from Italy. *Journal of Economic Geography*, 20(2):571–600.

Beine, M., Docquier, F., and Ozden, C. (2010a). Diaspora effects in international migration: key questions and methodological issues. Dem discussion paper series, Department of Economics at the University of Luxembourg.

Beine, M., Docquier, F., and Rapoport, H. (2008). Brain drain and human capital formation in developing countries: Winners and losers*. *The Economic Journal*, 118(528):631–652.

Beine, M., Docquier, F., and Rapoport, H. (2010b). On the robustness of brain gain estimates. *Annals of Economics and Statistics*, (97/98):143–165.

Beine, M., Docquier, F., and Çağlar Özden (2011). Diasporas. *Journal of Development Economics*, 95(1):30 – 41. Symposium on Globalization and Brain Drain.

Beine, M., Docquier, F., and Özden, (2015). Dissecting network externalities in international migration. *Journal of Demographic Economics*, 81(4):379–408.

Beine, M., Noël, R., and Ragot, L. (2014). Determinants of the international mobility of students. *Economics of Education Review*, 41:40–54.

Beine, M. and Parsons, C. (2015). Climatic factors as determinants of international migration. *The Scandinavian Journal of Economics*, 117(2):723–767.

Belot, M. and Ederveen, S. (2012). Cultural barriers in migration between oecd countries. *Journal of Population Economics*, 25(3):1077–1105.

Bengtsson, L., Lu, X., Thorson, A., Garfield, R., and von Schreeb, J. (2011). Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in haiti. *PLoS medicine*, 8(8).

Berger, J. and Pope, D. (2011). Can losing lead to winning? *Management Science*, 57(5):817–827.

Bernard, T., Dercon, S., Orkin, K., and Taffesse, A. S. (2014). The future in mind: Aspirations and forward-looking behaviour in rural ethiopia. CSAE Working Paper Series 201416, Centre for the Study of African Economies, University of Oxford.

Bernard, T., Dercon, S., Orkin, K., and Taffesse, A. S. (2019). Parental aspirations for children's education: Is there a "girl effect"? experimental evidence from rural ethiopia. *AEA Papers and Proceedings*, 109:127–32.

Bertoli, S. and Fernández-Huertas-Moraga, J. (2012). Visa policies, networks and the cliff at the border. IZA Discussion Papers 7094, Institute of Labor Economics (IZA).

Bertoli, S. and Fernández-Huertas-Moraga, J. (2013). Multilateral resistance to migration. *Journal of Development Economics*, 102:79 – 100. Migration and Development.

Bertoli, S. and Fernández-Huertas-Moraga, J. (2015). The size of the cliff at the border. *Regional Science and Urban Economics*, 51:1 – 6.

Bertoli, S. and Ruyssen, I. (2018). Networks and migrants' intended destination. *Journal of Economic Geography*, 18(4):705–728.

Bisin, A. and Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298 – 319.

Blumenstock, J. E. (2012). Inferring patterns of internal migration from mobile phone call records: evidence from rwanda. *Information Technology for Development*, 18(2):107–125.

Boguslaw, J. (2017). *When the Kids Are Not Alright: Essays on Childhood Disadvantage and Its Consequences*. PhD thesis, Stockholm University, Faculty of Social Sciences, Department of Economics.

Borjas, G. J. (1992). Ethnic capital and intergenerational mobility. *The Quarterly Journal of Economics*, 107(1):123–150.

Bound, J. and Turner, S. (2002). Going to war and going to college: Did world war ii and the g.i. bill increase educational attainment for returning veterans? *Journal of labor economics*, 20(4):784–815.

Böhme, M. (2015). Migration and educational aspirations – Another channel of brain gain? *IZA Journal of Migration and Development*, 4(1):1–24.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Card, D. (2005). Is the new immigration really so bad?*. *The Economic Journal*, 115(507):F300–F323.

Card, D., DiNardo, J., and Estes, E. (2000). The more things change: Immigrants and the children of immigrants in the 1940s, the 1970s, and the 1990s. In *Issues in the Economics of Immigration*, NBER Chapters, pages 227–270. National Bureau of Economic Research, Inc.

Carlana, M., Ferrara, E. L. F. L., and Pinotti, P. (2018). Goals and gaps: Educational careers of immigrant children. Technical Report 1812, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.

Cavalli-Sforza, L., Menozzi, P., and Piazza, A. (1994). *The History and Geography of Human Gene*, volume 54.

Chiswick, B. R. and DebBurman, N. (2004). Educational attainment: analysis by immigrant generation. *Economics of Education Review*, 23(4):361 – 379. Special Issue In Honor of Lewis C. Solman.

Christofides, L. N., Hoy, M., Milla, J., and Stengos, T. (2015). Grades, aspirations, and post-secondary education outcomes. *Canadian Journal of Higher Education - Revue canadienne d'enseignement supérieur*, 45(1):48–52.

Coleman, M. and DeLeire, T. (2003). An economic model of locus of control and the human capital investment decision. *The Journal of Human Resources*, 38(3):701–721.

Collier, P. (2013). *How Migration Is Changing Our World*, volume 1. New York: Oxford University Press.

Collier, P. and Hoeffler, A. (2018). Migration, diasporas and culture: An empirical investigation. *Kyklos*, 71(1):86–109.

Comola, M. and Mendola, M. (2015). Formation of migrant networks. *The Scandinavian Journal of Economics*, 117(2):592–618.

Cornwell, C., Mustard, D., and Sridhar, D. J. (2006). The enrollment effects of merit-based financial aid: Evidence from georgia's hope program. *Journal of Labor Economics*, 24(4):761–786.

Cortes, K. E. (2006). The effects of age at arrival and enclave schools on the academic performance of immigrant children. *Economics of Education Review*, 25(2):121 – 132.

Czaika, M., de Haas, H., and Villares-Varela, M. (2018). The global evolution of travel visa regimes. *Population and Development Review*, 44(3):589–622.

Dalton, P. S., Ghosal, S., and Mani, A. (2016). Poverty and aspirations failure. *The Economic Journal*, 126(590):165–188.

DEMIG (2015a). Demig c2c, version 1.2, limited online edition. *"Oxford: International Migration Institute, University of Oxford"*.

DEMIG (2015b). Demig visa, version 1.4, full edition. *"Oxford: International Migration Institute, University of Oxford"*.

Deming, D. J. and Dynarski, S. (2010). *Into College, Out of Poverty? Policies to Increase the Postsecondary Attainment of the Poor*. University of Chicago Press.

DeMoss, K. (2013). Educational aspirations/expectations. In *Sociology of education: An a-to-z guide*, volume 1, pages 222–224. Thousand Oaks, CA: SAGE Publications, Inc.

Desmet, K., Breton, M. L., Ortuño-Ortín, I., and Weber, S. (2011). The stability and breakup of nations: a quantitative analysis. *Journal of Economic Growth*, 16(3):183–213.

Di Pietro, G. (2012). Does studying abroad cause international labor mobility? evidence from italy. *Economics Letters*, 117(3):632–635.

Di Pietro, G. (2015). Do study abroad programs enhance the employability of graduates? *Education Finance and Policy*, 10(2):223–243.

Di Pietro, G. (2019). University study abroad and graduates' employability. *IZA World of Labor*.

DNP (2015). Declaración de importancia estratégica del proyecto de apoyo a la formación de capital humano altamente calificado en el exterior. Technical report, Departamento Nacional de Planeación Bogotá, D.C., Colombia.

DNP (2016). Modificación al documento conpes 3835. declaración de importancia estratégica del proyecto de apoyo a la formación del capital humano altamente calificado en el exterior. Technical report, Departamento Nacional de Planeación Bogotá, D.C., Colombia.

Docquier, F., Peri, G., and Ruyssen, I. (2014). The cross-country determinants of potential and actual migration. *International Migration Review*, 48(1_suppl):37–99.

Docquier, F., Tansel, A., and Turati, R. (2020). Do emigrants self-select along cultural traits? evidence from the mena countries. *International Migration Review*, 54(2):388–422.

Docquier, F., Turati, R., Valette, J., and Vasilakis, C. (2019). Birthplace diversity and economic growth: evidence from the US states in the Post-World War II period. *Journal of Economic Geography*, 20(2):321–354.

Dustmann, C. and Glitz, A. (2011). Migration and education. In Hanushek, E. A., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 4, chapter 4, pages 327–204. North Holland.

Dustmann, C., Lanzara, G., and Frattini, T. (2012). Educational achievement of second-generation immigrants: an international comparison. *Economic Policy*, 27.

Dustmann, C. and Theodoropoulos, N. (2010). Ethnic minority immigrants and their children in britain. *Oxford Economic Papers*, 62(2):209–233.

Dynarski, S. (2000). Hope for whom? financial aid for the middle class and its impact on college attendance. *National Tax Journal*, 53(3.2):629–661.

Dynarski, S. (2003a). Loans, liquidity and schooling decision. Technical report, National Bureau of Economic Research.

Dynarski, S. (2005). Building the stock of college-educated labor. Working Paper 11604, National Bureau of Economic Research.

Dynarski, S. (2008). Building the stock of college-educated labor. *The Journal of Human Resources*, 43(3):576–610.

Dynarski, S. M. (2003b). Does aid matter? measuring the effect of student aid on college attendance and completion. *American Economic Review*, 93(1):279–288.

d'Hombres, B. and Schnepf, S. (2021). International mobility of students in italy and the uk: does it pay off and for whom? *Higher Education*.

Emonds, V. and van Tubergen, F. (2015). Mixed parents, mixed results: Testing the effects of cross-nativity partnership on children's educational attainment. *Sociological Perspectives*, 58(2):145–167.

Entorf, H. and Minoiu, N. (2005). What a difference immigration policy makes: A comparison of pisa scores in europe and traditional countries of immigration. *German Economic Review*, 6(3):355–376.

Faini, R. (2006). Remittances and the brain drain. *IZA Discussion Paper No. 2155*.

Feliciano, C. (2005). Does selective migration matter? explaining ethnic disparities in educational attainment among immigrants' children. *The International Migration Review*, 39(4):841–871.

Feliciano, C. and Lanuza, Y. R. (2017). An immigrant paradox? contextual attainment and intergenerational educational mobility. *American Sociological Review*, 82(1):211–241.

Fielden, J., Middlehurst, R., and Woodfield, S. (2007). Global horizons for uk students: a guide for universities. (project report)s. Technical report, London, U.K. : Council for Industry and Higher Education.

Figlio, D., Giuliano, P., Özek, U., and Sapienza, P. (2019). Long-term orientation and educational performance. *American Economic Journal: Economic Policy*, 11(4):272–309.

Fiorio, L., Abel, G., Cai, J., Zagheni, E., Weber, I., and Vinué, G. (2017). Using twitter data to estimate the relationship between short-term mobility and long-term migration.

In *Proceedings of the 2017 ACM on Web Science Conference*, WebSci '17, page 103–110, New York, NY, USA. Association for Computing Machinery.

Gang, I. N. and Zimmermann, K. F. (2000). Is child like parent? educational attainment and ethnic origin. *The Journal of Human Resources*, 35(3):550–569.

Garip, F. and Asad, A. L. (2016). Network effects in mexico–u.s. migration: Disentangling the underlying social mechanisms. *American Behavioral Scientist*, 60(10):1168–1193.

Gelman, A. and Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3):447–456.

Genicot, G. and Ray, D. (2017). Aspirations and inequality. *Econometrica*, 85(2):489–519.

Gibson, J. and McKenzie, D. (2011). Eight questions about brain drain. *Journal of Economic Perspectives*, 25(3):107–28.

Giuliano, P., Spilimbergo, A., and Tonon, G. (2013). Genetic distance, transportation costs, and trade. *Journal of Economic Geography*, 14(1):179–198.

Giulietti, C., Wahba, J., and Zenou, Y. (2018). Strong versus weak ties in migration. *European Economic Review*, 104:111 – 137.

Goux, D., Gurgand, M., and Maurin, E. (2017). Adjusting your dreams? high school plans and dropout behaviour. *The Economic Journal*, 127(602):1025–1046.

Grogger, J. and Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, 95(1):42 – 57. Symposium on Globalization and Brain Drain.

Head, K. and Mayer, T. (2014). Gravity equations: Workhorse,toolkit, and cookbook. volume 4, chapter Chapter 3, pages 131–195. Elsevier.

Hunt, J. (2011). Which immigrants are most innovative and entrepreneurial? distinctions by entry visa. *Journal of Labor Economics*, 29(3):417–457.

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635. The regression discontinuity design: Theory and applications.

Jacob, B. A. and Wilder, T. (2010). Educational expectations and attainment. Working Paper 15683, National Bureau of Economic Research.

Jaeger, D. A., Ruist, J., and Stuhler, J. (2018). Shift-share instruments and the impact of immigration. Working Paper Series 24285, National Bureau of Economic Research.

Jung, J. and Lee, Y. (2019). Aspirations versus expectations: Performance feedback, analyst forecasts, and corporate downsizing. *Academy of Management Proceedings*, 2019(1):15009.

Kane, T. J. (2007). Evaluating the impact of the d.c. tuition assistance grant program. *The Journal of Human Resources*, 42(3):555–582.

Kao, G. and Tienda, M. (1995). Optimism and achievement: The educational performance of immigrant youth. *Social Science Quarterly*, 76(5):1–19.

Kosec, K., Akramov, K. T., Mirkasimov, B., Song, J., Kamiljon, Akramov, and Song, J. (2018). Aspirations and women's empowerment evidence from kyrgyzstan. Ifpri discussion paper 01786, The International Food Policy Research Institute.

Krieger, T., Renner, L., and Ruhose, J. (2018). Long-term relatedness between countries and international migrant selection. *Journal of International Economics*, 113:35 – 54.

Latham, G. P. and Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes*, 50(2):212 – 247. Theories of Cognitive Self-Regulation.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

Lent, R. W., Brown, S. D., and Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, 45(1):79 – 122.

Levels, M., Dronkers, J., and Kraaykamp, G. (2008). Immigrant children's educational achievement in western countries: Origin, destination, and community effects on mathematical performance. *American Sociological Review*, 73(5):835–853.

Lochmann, A., Rapoport, H., and Speciale, B. (2019). The effect of language training on immigrants' economic integration: Empirical evidence from france. *European Economic Review*, 113:265–296.

Lu, X., Wrathall, D. J., Sundsøy, P. R., Nadiruzzaman, M., Wetter, E., Iqbal, A., Qureshi, T., Tatem, A., Canright, G., Engø-Monsen, K., and Bengtsson, L. (2016). Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in bangladesh. *Global Environmental Change*, 38:1–7.

Mahajan, P. and Yang, D. (2017). Taken by storm: Hurricanes, migration networks and u.s immigration. *NBER Working Paper Series, National Bureau of Economic Research*, (23756).

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

Massey, D. S. and Espinosa, K. E. (1997). What's driving mexico-u.s. migration? a theoretical, empirical, and policy analysis. *American Journal of Sociology*, 102(4):939–999.

Mau, S., Gülzau, F., Laube, L., and Zaun, N. (2015). The global mobility divide: How visa policies have evolved over time. *Journal of Ethnic and Migration Studies*, 41(8):1192–1213.

McKenzie, D. and Rapoport, H. (2010). Self-selection patterns in mexico-u.s. migration: The role of migration networks. *The Review of Economics and Statistics*, 92(4):811–821.

Mocan, N. and Tekin, E. (2011). Obesity, self-esteem and wages. In *Economic Aspects of Obesity*, NBER Chapters, pages 349–380. National Bureau of Economic Research, Inc.

Mocan, N. H. and Yu, H. (2017). Can superstition create a self-fulfilling prophecy? school outcomes of dragon children of china. Working Paper 23709, National Bureau of Economic Research.

Morgenstern, O. (1935). Volkommene voraussicht und wirtschaftliches gleichgewicht. *Zeitschrift für Nationalökonomie*, 6:337–57. Translated in English as "Perfect foresight and economic equilibrium", in Schotter 1976, 169-83.

Mountford, A. (1997). Can a brain drain be good for growth in the source economy? *Journal of Development Economics*, 53(2):287–303.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the u. s. labor market. *The Quarterly Journal of Economics*, 118(2):549–599.

Munshi, K. (2016). *Community Networks and Migration*, pages 630–648. Oxford University Press, London.

OECD (2011). *Education at a Glance 2011: OECD Indicators*. OECD Publishing.

OECD (2020a). *Education at a Glance 2020: OECD Indicators*. OECD Publishing.

OECD (2020b). *International Migration Outlook 2020*. OECD Publishing.

Oosterbeek, H. and Webbink, D. (2011). Does studying abroad induce a brain drain? *Economica*, 78(310):347–366.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Ours, J. C. and Veenman, J. (2003). J. the educational attainment of second-generation immigrants in the netherland. *Journal of Population Economics*, 16:739–753.

Ozden, C., Parsons, C. R., Schiff, M., and Walmsley, T. L. (2011). Where on earth is everybody ? the evolution of global bilateral migration 1960-2000. Policy Research Working Paper Series 5709, The World Bank.

Palacios, N., Guttmannova, K., and Chase-Lansdale, P. L. (2008). Early reading achievement of children in immigrant families: Is there an immigrant paradox? *Developmental Psychology*, 44(5):1381–1395.

Parey, M. and Waldinger, F. (2011). Studying abroad and the effect on international labour market mobility: Evidence from the introduction of erasmus. *The Economic Journal*, 121(551):194–222.

Pedersen, P., Pytlikova, M., and Smith, N. (2008). Selection and network effects—migration flows into oecd countries 1990–2000. *European Economic Review*, 52(7):1160–1186.

Pemberton, T. J., DeGiorgio, M., and Rosenberg, N. (2013). Population structure in a comprehensive genomic data set on human microsatellite variation. *G3 (Bethesda)*, 3(5):891–907.

Pinotti, P. (2017). Clicking on heaven's door: The effect of immigrant legalization on crime. *American Economic Review*, 107(1):138–68.

Portes, A. and MacLeod, D. (1999). Educating the second generation: Determinants of academic achievement among children of immigrants in the united states. *Journal of Ethnic and Migration Studies*, 25(3):373–396.

Portes, A. and Rumbaut, R. G. (1996). *Immigrant America : A Portrait*. A centennial book. University of California Press.

Quaglia, R. J. and Cobb, C. D. (1996). oward a theory of student aspirations. *Journal of Research in Rural Education*, 12:127–132.

Reyes, S. (1995). *"Educational Opportunities and Outcomes: The Role of the Guaranteed Student Loan*. PhD thesis, Harvard University. unpublished thesis.

Richerson, P. and Boyd, R. (2008). *Not By Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press.

Riphahn, R. T. (2003). Cohort effects in the educational attainment of second generation immigrants in germany: An analysis of census data. *Journal of Population Economics*, 16(4):711–737.

Ross, P. H. (2019). Occupation aspirations, education investment, and cognitive outcomes: Evidence from indian adolescents. *World Development*, 123:104613.

Rumbaut, R. G. (2004). Ages, life stages, and generational cohorts: Decomposing the immigrant first and second generations in the united states. *International Migration Review*, 38(3):1160–1205. An optional note.

Salisbury, M. H., An, B. P., and Pascarella, E. T. (2013). The effect of study abroad on intercultural competence among undergraduate college students. *Journal of Student Affairs Research and Practice*, 50(1):1–20.

Santos Silva, J. and Windmeijer, F. (1997). Endogeneity in count data models: An application to demand for health care. *Journal of Applied Econometrics*, 12:281–94.

Santos Silva, J. M. C. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Schnepf, S. V. (2007). Immigrants' educational disadvantage: an examination across ten countries and three surveys. *Journal of Population Economics*, 20(3):527–545.

Scott-Clayton, J. (2011). On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *The Journal of Human Resources*, 46(3):614–646.

Sewell, W. H., Haller, A. O., and Portes, A. (1969). The educational and early occupational attainment process. *American Sociological Review*, 34(1):82–92.

Smith, J. P. (2003). Assimilation across the latino generations. *American Economic Review*, 93(2):315–319.

Sorrenti, G. (2017). The spanish or the german apartment? study abroad and the acquisition of permanent skills. *Economics of Education Review*, 60:142–158.

Spilimbergo, A. (2009). Democracy and foreign education. *American Economic Review*, 99(1):528–43.

Spolaore, E. and Wacziarg, R. (2009). The Diffusion of Development*. *The Quarterly Journal of Economics*, 124(2):469–529.

Spolaore, E. and Wacziarg, R. (2016). *Ancestry, Language and Culture*, pages 174–211. Palgrave Macmillan, London.

Spolaore, E. and Wacziarg, R. (2018). Ancestry and development: New evidence. *Journal of Applied Econometrics*, 33(5):748–762.

Stark, O., Helmenstein, C., and Prskawetz, A. (1997). A brain gain with a brain drain. *Economics Letters*, 55(2):227–234.

Stark, O., Helmenstein, C., and Prskawetz, A. (1998). Human capital depletion, human capital formation, and migration: a blessing or a "curse"? *Economics Letters*, 60(3):363–367.

State, B., Rodriguez, M., Helbing, D., and Zagheni, E. (2014). *Social Informatics. SocInfo 2014. Lecture Notes in Computer Science*, chapter Migration of Professionals to the U.S, pages 531–543. Springer, Cham.

Stuen, E. T., Mobarak, A. M., and Maskus, K. E. (2012). Skilled immigration and innovation: Evidence from enrolment fluctuations in us doctoral programmes. *The Economic Journal*, 122(565):1143–1176.

Sweetman, A. and van Ours, J. C. (2014). Immigration: What about the Children and Grandchildren? IZA Discussion Papers 7919, Institute of Labor Economics (IZA).

Teichler, U. (2004). Temporary study abroad: the life of erasmus students. *European Journal of Education*, 39(4):395–408.

Thissen, L. and Ederveen, S. (2006). Higher education; time for coordination on a European level? CPB Discussion Paper 68.rdf, CPB Netherlands Bureau for Economic Policy Analysis.

Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6):309–317.

Tjaden, J. D. and Hunkler, C. (2017). The optimism trap: Migrants' educational choices in stratified education systems. *Social Science Research*, 67:213–228.

Tjaden, J. D. and Scharenberg, K. (2017). Ethnic choice effects at the transition into upper-secondary education in switzerland. *Acta Sociologica*, 60(4):309–324.

United Nations, D. o. E. and Social Affairs, P. D. (2015). International migration flows to and from selected countries: The 2015 revision. (POP/DB/MIG/Flow/Rev.2015).

Van Der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach*. *International Economic Review*, 43(4):1249–1287.

Vidal, J.-P. (1998). The effect of emigration on human capital formation. *Journal of Population Economics*, 11(4):589–600.

Von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Whatley, M. (2017). Financing study abroad: An exploration of the influence of financial factors on student study abroad patterns. *Journal of Studies in International Education*, 21(5):431–449.

Whatley, M. (2019). Study Abroad Participation: An Unintended Consequence of State Merit-Aid Programs? *Research in Higher Education*, 60(7):905–930.

Wilson, R., Zu Erbach-Schoenberg, E., Albert, M., Power, D., Tudge, S., Gonzalez, M., Guthrie, S., Chamberlain, H., Brooks, C., Hughes, C., Pitonakova, L., Buckee, C., Lu, X., Wetter, E., Tatem, A., and Bengtsson, L. (2016). Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 nepal earthquake. *PLoS Currents*.

Winkelmann, R. (2008). *Econometric Analysis of Count Data*, volume 1. Springer-Verlag Berlin Heidelberg.

Winters, P., de Janvry, A., and Sadoulet, E. (2001). Family and community networks in mexico-u.s. migration. *The Journal of Human Resources*, 36(1):159–184.

Zagheni, E., Garimella, V. R. K., Weber, I., and State, B. (2014). Inferring international and internal migration patterns from twitter data. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14 Companion, page 439–444, New York, NY, USA. Association for Computing Machinery.

Zagheni, E. and Weber, I. (2012). You are where you e-mail: Using e-mail data to estimate international migration rates. In *Proceedings of the 4th Annual ACM Web Science Conference*, WebSci '12, page 348–351, New York, NY, USA. Association for Computing Machinery.

Zealand, S. N. (2012). Using cellphone data to measure population movements: Experimental analysing following the 22 february 2011 christchurch earthquake. Technical report, Wellington: Statistics New Zealand.

Zuppann, A. and Liu, E. (2016). Relative obesity and the formation of non-cognitive abilities during adolescence. unpublished.

Zwick, R. and Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and sat scores: The role of student ethnicity and first language. *American Educational Research Journal*, 42(3):439–464.