Essays in applied economics

Martín Fernández Sánchez

UNIVERSITÉ PARIS 1 PANTHÉON-SORBONNE
ÉCOLE D'ÉCONOMIE DE PARIS

Pour l'obtention du grade de Docteur en Sciences Économiques par l'Université Paris 1 Panthéon-Sorbonne et l'École d'Économie de Paris.

Présentée et soutenue publiquement le 18 novembre 2020 par Martín Fernández Sánchez.

Thèse :

Essais en Économie Appliquée : Émigration, Identité et Bien-être

Martín Fernández Sánchez

Composition du Jury

Directeurs de thèse :

      Andrew Clark (Paris School of Economics - CNRS)

      Hillel Rapoport (Paris School of Economics - Paris 1 Panthéon Sorbonne)

Présidente :   Ekaterina Zhuravskaya (Paris School of Economics - EHESS)
Rapporteur :   Paola Giuliano (UCLA)
Rapporteur :   Paolo Pinotti (Bocconi University)
Examinatrice :   Libertad González (Pompeu Fabra University)

Acknowledgements

This thesis is the end of a long journey that I recklessly started several years ago. It has not been, by any means, an easy trip. Yet, it has been a pilgrimage of incredible personal growth, an extremely rewarding experience, and of course, a chapter of my life of which I will always keep fond memories. It represents, at the same time, the starting point of a new adventure and a stepping stone for more ambitious personal and professional goals.

Who could have imagined that that restless countryside boy who spent his time staring at bugs and daydreaming about discovering dinosaurs remains would go abroad and eventually do a PhD in economics in one of the best universities in the world? Frankly, not me until a few years ago. Of course, not my parents, who none of them had attended college and always lived less than 50 kilometers away from where they were born. Nobody, absolutely nobody, could have remotely conceived such a foolish idea ten years ago, but it has happened. Only by understanding the origins, the obstacles and whims of fate, as well as the tremendous help and support received along the way, it becomes evident that there is lot to be proud of, but more importantly, to be humble and grateful.

In the following lines, I will sketch how this came to happen, and while I tell such story, I will try to thank and praise all of those who, one way or another, made it happen, or contributed to make it a better story to tell. It will be an exercise of self-reflection and improvisation, as all I write hereafter will be spontaneous, in the absence of a plan or script, and it will not be changed afterwards, in fact, it will not even be re-read. My apologies hence for all the grammatical mistakes you may encounter. It will be a chaotic tale, full of time leaps, incoherent passages and digressions, but that is what makes a great fable. It may also contain a small dose of sentimentalism, but that is Murakami to blame (or to praise), whom I have been reading assiduously in the past few months.

Since most stories start from the beginning, I will start from the end and go in reverse, as in *The Curious Case of Benjamin Button*. This implies beginning by talking about my advisors and my thesis jury, which will be the last people my blue gaze will pierce before (hopefully) becoming a doctor. Doctor... that word still sounds too unfamiliar. I would like to thank my advisors, Andrew Clark and Hillel Rapoport, for taking me under their wings and make this PhD a reality. I will always be thankful for all the opportunities you selflessly offered me. Thank you for supporting each of my ideas no matter how harebrained they were, and for easing my fears in the moments of highest despair. Thank you for inspiring me with your own work and for defining the researcher I am today. Andrew,

you agreed to supervised the master dissertation of a student whose conversation you could probably barely work out, and you opened the Pandora's box of the PhD. Kudos to you. Sorry for making your eyes (and ears) bleed with my poor English, you have made it improve much more than you think. Hillel, you fearfully decided to join this odyssey, which was key to obtain a PhD scholarship, and you planted the seed of doing a stay in the US. Kudos to you. I hope that the end of my PhD is just the start of a long-lasting friendship and I dream about the day you will both finally visit the green land of Galicia. By the way, shall I not succeed in academia, I will try to convince you to form a rock band which I would anticipate have an unprecedented success. Actually, who says we cannot do both...

I need to thank all the members of my jury: Ekaterina Zhuravskaya, Paola Giuliano, Paolo Pinotti, and Libertad González. It has been an honor and a great privilege to have such an outstanding group scholars examining my work and helping me to improve it. I felt like in one of this TV shows in which an amateur singer is coached by a professional and well-known one. Well, I had the luck to have a whole gang. Thank you so much for all your feedback and advice, you have been terrific. I hope we keep in touch and we can meet each other soon. I am also looking forward to having the chance to collaborate with you in exciting projects in the future. Of course, you are also invited to visit Galicia whenever you want. These are not words to be gone with the wind, I truly mean it.

My next thoughts come to the people I met during my research stay at Harvard University in 2018 (and briefly again in 2019). Yes, you heard well, Harvard University. I still have to pinch myself to believe I was there. I was incredibly fortunate to have Alberto Alesina and Nathan Nunn inviting me to visit the economics department. For those of you who are not economists, this was the equivalent of meeting and hanging out with two celebrities you deeply admire, think as if *The Rolling Stones* invited you to tour with them. Their work has been a true inspiration and made a tremendous influence on me. Attending their courses, the seminars they organized, and discussing with them, was as close to a dream as it can get. Alberto, wherever your soul is (I bet in a snowy summit), thank you so much for having taught me a few invaluable lessons. Nathan, thank you so much for having shaped my interests and the way I see and think about the world. During my visit, I was also very lucky to meet and work closely with Sam Bazzi from Boston University. He integrated me in his group of students, treated me as an offspring, and supported me enormously. I miss my Tuesday ride across the Charles River to attend the presentations and discussions of the Development Reading Group, I will not miss the Domino's pizza though. I would also like to spend a few words thanking all the professors from both Harvard and Boston University I had the pleasure to meet, the people that made possible that I presented my work and tell the world about the Galician diaspora, and finally, all the students, seminar presenters and participants from whom I learnt a great deal. I would also like to thank Paris 1 University and the Académie Française for funding this dream. Without doubt, I had a wonderful time at Harvard, probably the most stimulating, prolific and hectic period during my entire PhD. I worked and learnt a lot, but I also had immense fun. I cannot close this paragraph without mentioning some of the amazing people I met there and with whom I built unbreakable bonds: Clemence, Lorenzo, Gemma, Eli, Laks...You made me laugh, enjoy life, and believe in myself, as nobody had done before. Thank you.

Let's talk now about what was my home between 2014 and 2020, with of course, several parenthesis during that period: the Paris School of Economics (PSE). Notice here how I talk about PSE and not Paris, because after these many years, I cannot separate both entities anymore. Let me first make a small digression and reveal something very few people are aware of. I was at the verge of not coming to PSE. Not because I did not want to, there was nothing I could have craved for more, but because I was not initially admitted. Indeed, I was placed 7th in the waiting list (of a 20 people program) and in June I was told the cohort was closed. As we would say in Spain, it felt like if somebody threw a bucket of cold water on me. I hence got ready for what would be my next destination: Barcelona. Paid the university fees, a deposit in a residence, readjusted my expectations... And then, in late August, after the insane Summer Carnival of my hometown, I accidentally opened my email box and saw an email from the PSE secretary sent a few days ago. One slot had opened up, and they needed an answer from me before Monday at 10am. Ehem... The next morning. I had not checked at all my email box for the past days, how could I be so lucky? I would lie if I say I did not cry a river. It was such a mixture of feelings. I did hesitate for a second, thinking about all the money already invested in the Barcelona plan and the mental effort I had made to imagine myself there. Yet, PSE was the school I wanted to go to the most to begin with so... But what can I tell you, you already know the ending.

My masters in Public Policy and Development in PSE (2014-16) was spectacular, if that is a word one can use to describe a masters. I probably never accumulated so much knowledge in such a short period of time. I enjoyed the classes, the work dynamics, I felt deeply inspired... Funnily enough, I went on from not being admitted into the program to becoming arguably one of the best students. Between my first and second year, I had the chance to do an internship at the OECD. There, I worked in the report of the Chilean economy and, as a result of that, I would later start thinking about what would become my dissertation topic and the third chapter of this thesis. I could go on for pages and pages talking about my adventures during the masters and these two years of *folie*, but I will just be concise and thank all the professors, many of them truly exceptional, as well as the outstanding secretary team that made the universal chaos of PSE (at least at the time) somehow operate. I also need to thank the firm ENCE, from Galcia, who generously gave me a scholarship to pursue this program. Of course, I met some very special people who managed to put me away from the books and with whom I shared unforgettable moments: Claire, Andrea, Adam, Stan... You were like distinct parts of a device that put together would create magic. I will always remember our trip to Reims, to the Lake District, the parties at Stan's place, our days at the ENS, as well as bicycling rashly around the streets of Paris (which almost brought me to my grave in a couple of occasions). I hope that we can keep collecting memories and funny anecdotes for many years. You had already been invited to Galicia, please come.

I remember the day the PhD scholarships were awarded as if it had been yesterday. It was in July 2016 and I was about to *"do the Americas"* for the first time and do an internship at the United Nations in Chile (I still do not know how they hired me without having an internal contact!). I had organized a dinner with some PSE colleagues to celebrate my departure and the end of the masters. Adam, Claire, Andrea, Stan, Pauline, Lucho, and

some others were there. We received an email notifying us about the scholarships, Pauline had been blessed, Lucho had not. The competition was harsh because there were very few grants to share between dozens of candidates from two master programs. I did not want to tarnish what should be a memorable *soirée* with either good or bad news. So I did not read the email. In fact, I did not read it until the next morning, and having my father on video call at the same time. You must know that there was no plan B, I had not applied to any other PhD program (because of what I will explain later), either I embarked in the PhD at PSE fully funded or I would reconsider my career after the internship in Chile. My dad did cried a river, it was the second time in my life I saw him crying (the other one being when my grandfather died). But what can I tell you, you already know the ending.

I came back from Chile in October 2016 with a bag full of fond memories and experiences, a marvelous summer love story, and crucially, unique data for my thesis. The first months of the PhD were a bit rough. They kept postponing the opening of the new PSE building. I spent several months without an office until Pepe and Julien kindly hosted me in theirs at the ramshackle Building E. Thank you guys! Similarly, I used to kill my time copying data from pictures into spread sheets, and wondering if that was what a PhD was about. I later realized that was what economic history is about when you do not have money to hire a research assistant. Things soon improved as I soon to frequent research seminars, not only to have a free lunch, but to see if I could learn something and also break the monotony those early days. I did not miss a single WIP seminar with the hope of being invited for a delicious meal at the Indian restaurant next to PSE. Such day, I would eat so much that I could barely work in the afternoon. Thank you Andrew for so many unproductive afternoons. I used to attend way too many seminars, sometimes I had to contain myself, but I admit that it gave me a lot of joy. Almost at the closing of my first year of the PhD, and after a fire that cause a short-circuit in the new building, I could settle in my brand new office which I would fill with junk over the course of the next three years. It is time to thank some of the people who shared those four walls with me and who took care of my beloved plants when I was away. Quitterie, Hélène, Celia, Thiago. . . Having you there felt like a ray of sunshine in the middle of a storm. Thank you. Coming back to the seminars, I also presented my research in PSE multiple times, 15 to be exact, what may constitute an all-time record. I would like to thank all the people who attended and who gave me advice. Similarly, all the professors that at some point or another helped me, in particular Oliver Vanden Eynde, who was part of my thesis committee. Thank you. Alike a rock star, I also toured extensively around Europe and the US presenting my research during my last two years. I would like to thank everybody who arranged those conferences and workshops and invited, making me feel like Persse McGarrigle in *Small World: An Academic Romance.*

The PhD experience improved significantly over time, as my understanding of what I was doing grew and as my network of friends and colleagues expanded. There is a large number of people who made my daily life at PSE much more enjoyable. Many names come to my mind such as: Pepe, Clara, Sofia, Yaz, Justine, Antton, Lucho, Mónica, Óscar, Paolo, Sara, Paul's, Laura, Luis, Alex, Alessandro, Mariona, Cyprien, Lisa, Lee, Celia, Thiago, Lydia, Resuf, Nitin, Florencia, Santiago, Javi Soria, Eric, IPP and J-PAL people. . . but there were many more. I am obliged to highlight a few central characters in this story.

Pepe, you have been like a bright start, a true mentor, a confidant, and an exceptional friend. I doubt this thesis would have been possible without your support and guidance. I will be eternally grateful. We had the time our lives when we both were in Paris. Our friendship also took us into some surrealist adventures in Murcia, Alicante, Kenya, Tanzania. . . I cannot wait for more. Clara, my job market buddy, my port amidst the worst tempests, you were the one who sustained me in the most difficult moments. I grew a lot working hand by hand with you, and you made me enjoy the job market experience (blasphemy!). You were the one motivating me to think about the Galician mass migration. I do not know how to thank you. I am missing already our football matches, but I am sure we will play again. Once again, thank you for everything. By the way, Clara is the only PSE person who has visited me in Galicia. What are you guys waiting for? I proceed with Yaz and Sofia. Thank you for so many special moments, laughs, support, and for irrupting in my life as two incredible friends. You have brought me too much happiness. You cannot imagine how much I miss our lunches and dinners at the Cité Universitaire. Sofia, thank you for all the interesting discussions we have had. You have broaden my horizons. Yaz, thank you for accepting the challenge of learning how to dance salsa together and for inspiring me to adopt great habits in my life.

I cannot but thank once again the Paris School of Economics as whole. I think it is a unique community and I feel extremely honored to have been part of it. Special thanks to Véronique Guillotin, the PhD secretary, who helped us in countless occasions, including the *voyage* through the wild waters of the job market. In fact, I would like to thank everyone who was involved in preparing us for such challenge and put her grain of salt. As I mentioned earlier, despite the similarities with *The Hunger Games* and probably being at the brink of depression at some moments, I really enjoyed the experience of the academic job market. When I was in Rotterdam and the US with all the other candidates, I felt I was taking part in some kind of tribal ritual. I sometimes imagined us as gladiators in a roman circus, with our job market papers as weaponry. Yet, I sometimes thought of myself as a *retiarius*, the gladiator with the net (seriously, a net against lions?!). Changing of arena, I would like to thank Marc Gurgan for giving me the opportunity to teach in the PPD master. It was a memorable experience which taught me a lot. Thanks to all my students for serving as guineas pigs. Thank you Thomas Piketty for allowing me organize the Applied Economics Lunch seminar and, as a result, always enjoy a tasty dessert. I am grateful also to the LABEX OSE who granted me a research fund that I could use to collect some data for my job market paper, to Pairs 1 Panthéon Sorbonne who granted me the PhD scholarship, and to the French Republic and the welfare system which allow me to do my fourth year while on unemployment benefits. In short, thanks to all of the residents of the 48 Boulevard Jourdan, a place which will always feel like home.

There were also many other people with a prominent role during my time in Paris, who contributed enormously to my happiness and who made me grow in many dimensions. I think, for instance, about all the people I met at the Cité Universitaire de Paris, were I was blessed to be hosted for four years. Among them, the Tahitian dance group I met at the Maison des Province de France, the guys from the Fondation Biermans-Lapôtre with whom I formed a jazz band and performed in several venues, Fernando with whom I participated in some amazing jam sessions and at the concert to celebrate the 10th anniversary of

PSE, many friends from the Colegio de España and the Portuguese House, etc. Thanks to Adrien, my coloc during my last year and half in Paris, with whom I had very interesting debates and funny anecdotes that I will keep preciously for the rest of my life. I would also like to mention Flor, whose artistic spirit fascinated me and whose golden heart impressed me even more. And finally, Adriana, with whom I shared many memorable sunsets and sunrises, discovered Madrid and the South of Spain, and rediscovered Galicia. Thanks to all of you for your part on this plot.

An intuitive reader will have understood this is approaching to its end, or its beginning, depending on how one looks at it. I would like to devote a couple of lines to tell you very briefly about the days prior to *The City of Lights*. First, I feel the need to go back to my teenage years. I did warn you about the time leaps. In high school I was not a brilliant student, if anything, I was a mediocre. Yet, I only failed one subject during student years, Mathematics (which is pure irony given that I became an economist). In my early adolescents I had some friends who, to put it simply, were not the best influence. For a series of events which I cannot recollect very well, I changed friends several times around those years. Luck played an important role steering me clear from dangerous people and making find other friends some of whom are still very close. It was only in the last two years before college that I started to hit the books a bit more and attain somewhat decent grades. I would like to thank my teachers at IES Mendiño, especially Carmen Villanueva (history and geography), Carmen Portabales (economics), and Esperanza (philosophy).

To be blunt, I did not have any vocation to study economics. It was a mere accident. Indeed, I was more inclined towards philosophy and arts, but we can say that my parents exert a strong influence (slash manipulation) on my major's choice. Strikingly, I found an unexpected interest in economics which would help me to perform superbly in the degree. It is time to recognize the merit of the teachers I had at the University of Vigo who did an outstanding job. They know I appreciate them a lot and I still try to visit them whenever I have the chance. I would like to thank Miguel and Damián, still very good friends today, for the good old days in college. Other main characters outside university who definitely contributed to my metamorphosis during those years were Beto, Juan, Paula and Laura. You cannot even imagine how much of an influence you have exerted on me and the many fond memories I keep of you. Now, if you had talked to me in my third year of college, I would have told you I had not idea of what I wanted to do after. My econometrics professor, Daniel Miles, the toughest I remember, one day came to talk to me about this and mentioned that I should consider studying abroad. At that point, such idea had never crossed my dizzy mind. From there, a series of events took place creating the perfect storm. I attended a coaching and motivational workshop who made me believe I could achieve anything I could aspire to. I obtained a grant from the Galician Government which allowed me to go to study English for a couple of months in the UK. Finally, with the support of my undergraduate thesis advisor Carlos Gradín, and other professors like Gustavo Bergantiños and Loli Garza, I applied to several universities and somehow got admitted into the Cambridge University. To make things even more unreal, I received a grant from the Fundación Ramón Areces that would cover all the expenses. Once again, the starts aligned in an unthinkable way. I feel extremely indebted to my professors at the University of Vigo and the Fundación for making possible this crucial transition.

My time in Cambridge in 2013-13 had both bright and dark sides. I guess there is a thin line between a dream and a nightmare and one has to be a careful tightrope walker. This chapter did not start as expected. . . The kick-off of the academic year was a very intense two-week math camp. It was meant to be a revision, now for me one third of the material was completely new. As you can imagine, these mathematical concepts were as easy to digest as oysters. We were told we would do a test simply to know where we stood. I scored 55/100. A few days later, I was asked to meet the program's director, Donald Robertson, who nicely offered me three options: (1) *"pack my bags and return home"*, (2) change program and do a masters say in development studies, or (3) pursue the M. Phil in Economics research and fail (which would imply reimbursing the funds of my Spanish grant, something I would not be able to face). I tried to explain myself and convince him that I would do my best to catch up, that I would seek help, or do anything it could take. He offered no help, being useless and demoralizing. Guess what? I chose option three.

Cambridge was a life changer experience. Think about it, it was my first time living abroad, I was being surrounded by genius, in a place with so much history behind. . . It will not come as a surprise that every often I asked myself if I truly belonged there or there had been some kind of misunderstanding. I found the program quite challenging because, despite the good will of my teachers in Vigo and my outstanding transcript, I was not prepared for the kind of training and educational system that was awaiting me in Cambridge. Given that it was a one-year program, I also had to apply to many PhD and master programs for the forthcoming year as well as different funding opportunities. Doing all this was so time consuming and stressful that I admit I was down in the dumps. Incidentally, all my applications to US PhD programs were rejected, which explains why two years later I couldn't bother to go through the same. Nevertheless, I worked very hard during the masters and passed all my exams. I also had enough time to meet fascinating people. I will mention the Eddies, the gang of friends I had at St. Edmunds College (in particular my classmate Alexey), Héctor, who told me about the PPD program at PSE, and with whom I shared struggles and moments of joy, and Theresa, a mysterious girl with intriguing stories, who got me interested in international organizations, and who would appear and disappear like a ghost. You were all crucial shaping my future steps and I am extremely grateful for that.

So we come to the final words, and to the cornerstone of this journey. My family. Nothing of this would have been possible without the unconditional love and support of my parents. It does not matter that they do not fully understand what I have been going through or that they cannot always give me advice. They did their best at helping me in the ways they could. Protecting my health and peace of mind, cheering me up with I needed it the most, playing down success and failure, and being proud of every tiny achievement. If I tell them it is an important milestone, they are always there for me. I do not take for granted any of this, I have been infinitely fortunate to have such parents. I cannot but be moved and in a sea of tears while I write these words. Thank you mum, thank you dad, you are extraordinary. You are the true heroes of this story. You have provided for your family by working day and night, by moving mountains and facing all types of challenges. This is your achievement, not mine.

I would also like to have a few words for my brother, Pablo, who shaped me in many positive ways while I was a kid and who was a role model. Finally, I need to mention my grandmother, Regina, whose stories of hardship (many of them probably unreal) always inspired me to push my boundaries and be modest. She dropped out of school at the age of 9 and she worked precariously in the fields most of her life. That is where we come from and that is why this PhD is such an impressive accomplishment. I will defend this thesis one day after her 99 birthday, so it will be my tribute to her.

In sum, I was a mischievous kid lucky enough to dodge rotten apples and get surrounded by nice peers, inherited amazing parents who supported me and transmitted me great values, and realized how essential education is thanks to passionate professors. I studied economics by accident. I worked really hard and a series of random circumstances inspired me to aim high. Some people believed in my potential and I was extremely lucky to receive several grants that funded my studies in Cambridge and my PhD at the Paris School of Economics. Finally, I met extraordinary people who helped me lots along the way. Thank you to all of your one last time. You have made this happen. Yet, it has not been bed of roses. As I explained earlier, not everyone had faith in me, and in several occasions, I was at the edge of the precipice. What a crude reminder of the major role of luck in our lives. What we can interpret as *success* or *failure* is often the result of a coin toss, of *Tyche*'s caprice. I guess the moral of this fable is that we should not focus our attention so much on what we achieve or the outcomes, but rather on the effort and passion we put in the process of pursuing our goals, and in enjoying every step we take in such path.

So here I am, about to turn 30, about to finish my PhD, lying in the grass in a park in Belval, staring at bugs and daydreaming about discovering dinosaurs remains. You will hear from me if I ever do.

Written in Belval on 19 October 2020.

Martín Fernández Sánchez

Contents

11

Introduction

Big questions and history as laboratory

"Much progress in the social sciences is due to the genius of those who have discovered, critically evaluated and compared [. . .] natural experiments in fields where a laboratory experiment is impractical, if not impossible." – Ernst Mayr (1997, p.29)

The most remarkable research in economics over the past decades has been characterized by a combination of addressing relevant questions with the most convincing methods. This has involved focusing on big issues, exploiting techniques or research designs that provide appropriate counterfactuals, and having a deep understanding of the contexts under study.

One of the aspects that has received more attention has revolved around the sources of economic development. Why are some countries richer than others? What can be done to promote economic growth and prosperity? In words of Robert Lucas (1998). *"[...] The consequences for human welfare involved in questions like these are simply staggering: once one starts to think about them, it is hard to think about anything else."* These sort of questions date back to classical economists such as Adam Smith (1776) and David Ricardo (1817) and received major contributions in the second half of the twentieth century (Solow, 1957; Mankiw, Romer and Weil, 1991). Despite a stunning progress of many developing countries in several dimensions such as investments in physical capital or access to health and education (Deaton, 2013), we still see large differences in income levels across the world today, which has spurred a vivid debate over the fundamental causes of development. To explain these persistent inequalities, some economists have pointed the role played by factors such as geography (Sachs and Warner, 1995; Diamond, 1998), institutions (North, 1990; Acemoglu, Robinson and Johnson, 2001), and more recently culture and its interplay with the former (Tabellini; 2010, Nunn; 2011).

In recent times, not only the scope of economic research has expanded, but it has also experienced a rapid change in its approach, placing a greater emphasis on credibly estimating causal effects. Today more than ever, economists try to carefully evaluate the data they exploit, construct meaningful counterfactuals and clearly state the assumptions behind their research design, as well as provide some evidence that can give support to such assumptions. This so-called "credibility revolution" (Angrist and Pishke, 2010), which started in labor and public economics in the early 1990s (Angrist, 1990, Angrist and Krueger, 1991),

rapidly spread across fields with great importance in development economics (see the seminal works of Duflo, 2001; Miguel and Kremer, 2004) and more recently being incorporated into areas with traditionally a less quantitative approach such as economic history (some prominent examples are Nunn, 2008; Dell, 2010).

This credibility revolution has come hand in hand with an extraordinary rise in the use of historical natural experiments within economics (Margo, 2020). The reason for this is that history provides an ideal laboratory with unique sources of variation that allow addressing big question in a casual way (Cantoni and Yuchtman, 2020). In fact, historical events often allow testing hypothesis or explore questions that would have been impossible to address in modern times or with contemporary data. Another advantage of adopting a historical and long-term perspective is that it allows sufficient time to pass and thus one can examine how different dynamics evolve or to take into account general equilibrium effects.

There has been an explosion of studies using innovative research designs and a variety of techniques while taking advantage of historical data. For instance, original instrumental variables such as the direction of winds and their importance for different sailing technologies as used by Feyrer and Sacerdote (2009) to study the impacts of colonialism or by Pascali (2017) to explore the impact of trade on development. Distance to a border in a regression discontinuity design to understand the long-term effects of extractive institutions (Dell, 2010) or the agglomeration economies (Ahlfeldt et al., 2015). Also shift-share instruments that leverage variation from both spatial and time-changing factors (Sequeira, Nunn and Quian, 2020).

Amidst the recent boom of impact evaluations and randomized control trials, some development economists are raising their voices calling into question the usefulness of such approaches without a thoughtful reflection on the context of the studies. First, it is often unclear whether the particularities of the setting may make the findings difficult to extrapolate to other settings, what is commonly known as "external validity" (Deaton and Cartwright, 2016). Secondly, even without wondering about the external validity, a poor understanding of the context limits the ability of the researcher to comprehensively interpret the results and their implications (Nunn, 2019). As Nunn (2019) puts it, without realizing the importance of the local context, researchers will have an incomplete or incorrect view of the research questions themselves, and a poor understanding of the reasons why some policies may succeed or fail. To tackle this, more context-specific knowledge is need, qualitative work, and collaboration with local scholars across fields.

This dissertation is no exception in that it tries to address broad questions in the development and political economy literature. Rather than exploring a single topic and various questions around it, each chapter explores a different question in what could be seen as a research paper on its own (hence the use of the word "paper" in the chapters). Similarly, although the approach is always empirical, the techniques used in each chapter are also different, and particular suited for the question and context under study. The following chapters will cover aspects such as how does emigration affect the communities of origin, can emotional shocks affect the feelings of national identity and influence political preferences, or what are the effects of school segregation on long-term well-being.

As previously exposed, this thesis follows the approach of carefully using the data and econometric techniques in order to establish causal relationships. In the chapters I implements wide range of econometric techniques such as shift-share instruments, even studies and cohort analysis, among others.

Chapters 1 and 3 take advantage of historical natural experiments (i.e., the Galician diaspora during 1900-1930 in Chapter 1 and a major educational reform during Chile's dictatorship in Chapter 3) while Chapter 2 relies on a quasi-experimental setting in which individuals are randomly exposed to unexpected shocks (i.e., a match loss of Football Club Barcelona, a standard bearer of Catalan identity). Both Chapters 1 and 3 rely on unique historical data collected for the purpose of this thesis. In a heroic task of data collection, I did several trips to Spain, in particular to Madrid and Santiago, where I gathered historical data from the Spanish statistics of migration to Latin America, from educational and population censuses and many other sources. During these visits, I not only took thousands of pictures of these materials, I also took the opportunity to visit some libraries and read different books, thesis, and manuscripts, that would have been otherwise inaccessible. In the same spirit, just before starting my PhD I spent several months in Chile searching for some data that most researchers deemed as "lost forever". After getting in touch with dozens of people I finally managed to find those "lost" records in an archive of the Ministry of Education in the outskirts of Santiago. There, I found data on schools matriculation for all Chilean municipalities during the 1980s what constitutes the basis for the analysis of the 1981 voucher reform. In both cases, I had to spend several months typing data by hand into spread shits in order to construct the data-sets that later allowed me to carry out the analyses.

Finally, it is essential to highlight that this thesis does not only deal with questions of the type "what causes what", but also aims to quantify such relationships and delve into the "how and why", investigating the mechanisms behind the core findings. To guide such analysis, it relies on a deep understanding of the context of each study, and on an extensive reading of the literature on other fields such as history and sociology. These combination of ambitious research questions, rigorous methods, and a multidisciplinary approach, are the distinctive traits of this dissertation.

Related literature and main contributions

Migration and human capital

One fundamental question in the development literature is how emigration affects the countries of origin. Does emigration foster or hinder economic development (Aggarwal, Demirgüç-Kunt, Pería, 2011)? Does it deplete or increase the stock of human capital (Beine, Docquier and Rapoport, 2008)? Can it contribute to change institutions and consolidate democracies (Barsbai et al., 2017)? With a growing number of migrants in the past decades and with prospects of increasing flows in the years to come (United Nations,

2020), today these aspects have become more relevant than ever.

A large body of work has focused on examining the trends and patterns of international migration (Docquier and Marfouk, 2006). Scholars soon realized that migrants tend to be positively selected in terms of education (i.e., with higher education than the average of the population), and as a result, their departure could reduce the stock of human capital at origin (Carrington and Detragiache, 1998). In face of this, some economists suggested that developing countries should implement a "brain tax" to compensate for the departure of these educated individuals (Bhagwati, 1976). It turns out that the positive selection of migrants is more the norm than the exception. In fact, more than 80% of all the countries seem to display such pattern today (World Bank, 2018) and is a phenomenon that has been increasing over time (Docquier and Rapoport, 2012). Furthermore, poorer countries tend to present higher emigration rates and a higher degree of positive selection (World Bank, 2018). The combination of these three facts seems to create a perfect storm, indicating that emigration could indeed be detrimental for the stock of human capital in the countries of origin, especially for the poorest ones.

Yet, some work around two decades ago already highlighted that the relationship between emigration and education is much more complex than simply a compositional effect. Beyond the direct effect caused by the departure of individuals with a certain level of education, the prospects to emigrate and remittances could also influence educational investments. Indeed, if the prospects to emigrate provide a sufficiently strong incentive to invest in human capital but only a limited number of these individuals eventually leave, the net impact of emigration on the stock of human capital would not need to be negative, and indeed, could even be positive (Beine, Docquier and Rapoport, 2001).

In recent years there has been a large number of studies examining how emigration affected the family left behind, and in particular, the educational outcomes of children. For instance, Antmann (2012) study the Mexican-US migration and suggests that emigration could be beneficial for the children left-behind, particularly for girls. She exploits differences in exposure to emigration based on the age of children when the parent migrates. Work by McKenzie & Rapoport (2006, 2010) in the same setting challenges this conclusion indicating that the impact of emigration may be more complex, and even depend on the migrant networks. In fact, if these networks allow individuals with low skill to earn higher wages at destination, it could induced children to leave school prematurely and thus have a negative impact on the stock of human capital.

These and similar studies faced the challenge that migration decisions are not exogenous. First, households with or without a migrant member tend to be different on a wide range of dimensions. Secondly, the timing at which a family member decides to emigrate may be correlated with other shocks and conditions affecting the household. Finally, in some instances entire households migrate making it more difficult to obtain data about them. Although authors have tried to implement research designs that provide a better counterfactual, in other words, what would have happened if such family had not migrated, or if there had not been emigration from that community, there remain concerns about the existence of confounding factors affecting migrant flows. To overcome these limitations,

other work has relied on rich surveys or quasi-experimental settings in which the option to emigrate is random. For instance, Batista, Lacuesta and Vicente (2011) take advantage of a rich household survey conducted in Cape Verde containing migration stories of all household members and estimate that the prospects of emigration increase individuals' educational attainment. Gibbons, McKenzie and Stillman (2011) exploit a lottery system to emigrate from Tonga to New Zealand to examine the impact of emigration on the family left-behind on a variety of outcomes, including education. They do not find any significant impact on literacy or years of education, but an important caveat is that in their setting the migrants are typically not the parents of the children for which they have data but rather their uncles and aunts. Shrestha (2017) examines the impact of introducing an educational requirement in the selection of Nepalese people to join British Army and finds large positive effects regardless of whether they were or not successful in their application. He documents thus that the establishment of such requirement led to net gains in human capital among the communities affected. While these studies provide important insights, the particularities of the settings studied and their research designs pose serious questions about the external validity of the findings. Moreover, they are generally agnostic about the mechanisms that could drive their effects.

Another important limitation of the previous literature is that it has tended to focus on the impact of migration on prospective migrants or the children left-behind around the time migration takes place. This short-term perspective is not ideal for a number of reasons. First, the children of migrants' may emigrate themselves, so without considering their future migration trajectories it is impossible to assess the net effect on the stock of human capital. Furthermore, a considerable share of migrants may return to their countries of origin, and hence, the magnitude of this inflow and the degree of selection into return migration can play a major role. Knowing when migrants tend to return is crucial to determine the time horizon that should be analyzed. A study that has adopted a more long-term perspective is Dinkelman and Mariotti (2016), which I discuss later.

The previous channels discussed imply changes in the composition of the population (i.e., who leaves, who stays, and who comes back), there are however several other mechanisms that could make the impact of emigration differ in the short and long run. For instance, migrants may invest in public goods in their home communities such schools or hospitals (Chauvet et al., 2014). These investments may in turn have a positive effect on children's educational achievement, and importantly, for the community as whole and not just the children of migrants. Yet, these investments may only take place after migrants have accumulated sufficient wealth or if a system of collective remittances is established, such as through migrant hometown associations. Similarly, migrants may engage in social remittances related to the value or returns education. By diffusing such norms or information they could shape individuals' decisions about how much human capital to acquire or invest on for their children. Yet, these processes of cultural change may take several generations to materialize. Importantly, in the long run emigration may imply all sort of general equilibrium effects that can only be captured if sufficient time has passed and if one looks at communities as a whole.

The first chapter of this thesis revisits the question of how emigration affects human capital accumulation at origin but it introduces several major innovations with respect to the previous literature. One of the main contributions is that it adopts a very long-term perspective and a historical approach. To do so, I focus on one of the greatest migration episodes in modern history, the Age of Mass Migration (1850- 1930) a period in which over 40 million Europeans crossed the Atlantic towards the Americas (Hatton and Williamson, 1998). In particular, I look at the Galician mass emigration to Latin America that took place between 1900 and 1930 and its effects on human capital accumulation thorough the rest of the century.

Leveraging new historical data with contemporary data spanning over one hundred years I document that emigration led to persistent gains in human capital in the communities of origin across several generations. Importantly, I implement a research design that exploits fluctuations in economic conditions at destination in an instrumental variables approach to obtain exogenous variation in the timing of emigration across Galician municipalities and a proxy of past migrant networks to infer the intensity of emigration. With this strategy and taking municipalities as the unit of analysis, I circumvent the issues of past research that had to deal with the selection of individuals or households into emigration.

This study thus contributes to the growing research exploring the long-run effects of migration. Recent work has documented positive long-run effects on economic development and human capital accumulation in the areas receiving migrants (Rocha, Ferraz and Soares, 2017; Droller, 2018; Sequeira, Nunn and Quian, 2020). Regarding the impacts at origin, the literature has found positive effects on innovation (Anderson, Karadja and Prawitz, 2020), and mixed results in terms of economic development and human capital formation (Dinkelman and Mariotti ,2016; Testa, 2020). I provide novel evidence that, despite a short-run loss, the Galician mass emigration led to gains in human capital that persisted over one hundred years. To understand the variety of results, it is key to understand the context of each study. Dinkelman and Mariotti (2016), for instance, examine the circular migration from Tanzania to South African mines, which for political reasons suddenly stopped forcing all migrants to return home. Given that all migrants had to return and that they received all their earnings upon arrival, this is a setting particularly beneficial for a positive impact on human capital formation. In the case of Testa (2020), he studies an episode of forced migration (Germans expelled from Czechoslovakia after WWII) meaning that a particular group of the population was expelled from their lands and not allowed to come back. He finds that long-run negative effect on human capital in the areas that experienced a larger exodus.

In this regard, the Galician diaspora offers a better context to analyze the effects of emigration at origin. First, at the beginning of the twentieth century Galicia shared many characteristics with other development countries today, such as an economy relying almost exclusively in agriculture and very low levels of human capital and educational infrastructure. Moreover, it was similar to other migration flows from Southern Europe during the age of Mass Migration and to other migration flows today, characterized by male migrants often leaving their family behind, and thus, maintaining strong links with their communities of origin in clear contrast with episodes of forced migration (Becker and Ferrara,

2019). Finally, the decisions to emigrate, return, or send remittances, were not influenced by institutional factors or wars and thus resembled those of most migration episodes today.

A final contribution of this work has to do with the mechanisms explored. I examine two novel channels through which emigration could affect human capital in the long-run, namely investments in public goods and cultural changes. Inspired by the work examining the role of non-governmental institutions financing public goods and contributing to local development (Valencia-Caicedo, 2019), this chapter investigates the contribution of migrants' associations. I document that Galician migrants established hundreds of associations with the aim of investing in the construction of schools in their home communities, what could be understood as a pioneer form of development aid. I analyze the factors influencing the origin of these associations, their decision to construct schools in their villages, and their impact on educational attainment and further migration.

This work also relates to a recent literature on migration and culture. For instance, Knudsen (2019) shows that migrants from Scandinavian countries were on average more individualistic and, as a result, their departure raised the level of collectivism at origin. Becker et al. (2020) show that Polish forced migrants subsequently became less materialistic and invested more in education given is portable nature. I contribute to this literature by examining how emigration affected beliefs about the value of education and effort. Rather than the result of migrant selection or forced migration (Becker et al., 2020), my findings seem to suggest that an exposure to an environment where education is highly valued may make migrants change their norms and transmit these to their communities of origin.

National identity and political preferences

Most countries in the word, especially developing ones, present high levels of diversity and strong identity cleavages. These may also come in the form of differences in language, religion, etc. Understanding the nature of such cleavages is of major importance as they may lead to conflict and hinder economic growth (Alesina and La Ferrara, 2005). Similarly, they can lead to distrusts among individuals from different groups, less cohesive societies and even under-investment in public goods (Guiso et al., 2016). Given the broad implications of ethnic cleavages, it does not come as a surprise that modern states have engaged in a variety of nation-building policies in order to foster a common sense of identity. These have include propaganda (Blouin and Mukand, 2019), mass schooling (Bandiera et al., 2013), and even population mixing (Bazzi et al., 2019). Yet, even if successful, the legacy of ethnic cleavages may persist for generations (Besley and Reynal-Querol, 2014).

A large body of literature has explored the economic consequences of ethnic diversity and segregation, which are at the heart of identity cleavages. For instance, Montalvo and Reynal-Querol (2005a, 2005b) show empirically that ethnic polarization can reduce economic development through their effect on the occurrence of conflict, by reducing investments and affecting government expenditures. A number of studies show that higher ethnic diversity tends to lead to lower investments in public goods (Banerjee, Lakshimi,

and Somanathan, 2005; Algan, Hémet and Laitin, 2016; Desmet, Gomes and Ortuño-Ortín, 2020). A reason for this is that more ethnically segregated societies tend to display lower levels inter-ethnic trust (Alesina and Zhuravskaya, 2011) and more difficulties to cooperate (Guiso et al., 2016). Recent research has highlighted that identities and cultural values interact in complex ways, and the way these two aspects overlap may have different economic and political implications (Desmet, Ortuño-Ortín and Wacziarg, 2017). Where do identity cleavages originate? High levels of ethnic diversity and polarization have often been the result of historical processes involving the settlement of populations and migration flows (Ahlerup and Olsson, 2012), geographical factors such as soil quality and elevation (Michalopoulos, 2012) and, of course, the conquests of territories, colonization, and the formation of new states integrating different ethnic groups. In the case of Africa, Michalopoulos and Papaioannou (2016) point to European colonization and the arbitrary drawing of boundaries as one of the main causes for ethnic polarization and conflict in the continent.

Another strand of the literature has delved into exploring the factors and policies that could foster a national identity. One of the dimensions that has received more attention is the educational system (Alesina, Giuliano and Reich, 2018). For instance, Bandiera et al. (2019) provide evidence that US states introduced compulsory schooling laws to spread civic values among the immigrant population. Cantoni et al. (2017) examine the impact of a curriculum reform in China, finding that it shaped political preferences but it did not have a meaningful effect on national identity or behavior. Clots-Figueras and Masella (2013) show that being exposed to a reform introducing compulsory Catalan in schools increase Catalan identification and preferences over self-government and secession in the region. There are also some cases of similar policies that were not successful or even had unintended consequences. For example, Chen, Lin and Yang (2018) analyzed the impacts of a curriculum reform in Taiwan finding that while it increased national identification, the effects disappeared over time. Fouka (2017) shows that the prohibition of German in US schools led to a reinforcement of their group identity among German immigrants.

Other the research has examined more radical nation-building policies such as massive population re-settlements aiming at mixing people from different origins (Bazzi et al., 2019). Some work has documented the powerful force of propaganda either promoting prejudice against particular ethnic groups (Voigtländer and Voth, 2015) or reducing feelings of own-group identity (Blouin and Mukand, 2019). Now, identity feelings and values can also be influenced by external threats or situations that drive emotional shocks, especially those involving collective experiences (Dell and Querubin, 2018; Depetris-Chauvin et al., 2020). For instance, Depetris-Chauvin et al. (2020) explore the impact of national sports successes on identity and conflict in the context of Sub-Saharan Africa. They find that victories of the national soccer team in international competitions can increase national identification (in detriment of ethnic identity), lead to higher inter-ethnic trust, and reduce the prevalence of conflict in the following months. They suggest that their effects are driven by the collective experience of major sporting events, although the particular channels are not fully understood.

The second chapter of this thesis builds on the previous work relating collective emotional

experiences and national identity, trying to shed some light on the mechanisms of such connection and its potential political consequences. To do so, it focuses on the independence movement in Catalonia and the influence of Football Club Barcelona (FCB) during the last decade as an ideal setting. In particular, it examines whether the performance of the club can have an impact on Catalan identity, political preferences, and ultimately electoral results. The analysis presented mimics an experiment in the sense that individuals are randomly exposed to a shock depending on the date they are interviewed. In other words, I compare individuals interviewed right before or after a game, depending on its outcome. With this methodology, I can pin down the causal impact of the emotional shock associated to FCB defeats/triumphs on individuals' attitudes.

Importantly, I not only look at identity but try to establish its bridge with political preferences and political outcomes under the same framework. To do so, I leverage data of all elections taking place in Catalonia during the democracy, which I combine with information on the matches. Again, I take advantage of the random nature of games played before elections to investigate their impact on electoral results. Connecting changes in identity with political outcomes is crucial because if there is a change in the politicians elected, these could in turn introduce policies or change institutions in a way that reinforce or change trends in identity formation and many other social and economic dimensions.

My work also contributes a large research on the consequences of sporting events and the performance of national and local teams. Previous research has shown that, across a variety of sports, the performance of local teams can affect crime (Munyo and Rossi, 2013), domestic violence (Card and Dahl, 2011), students' achievement (Lindo et al., 2012), judicial decisions (Erenand Mocan, 2018) and even infant health (Duncan et al., 2017). Some of these studies has suggested that emotions can trigger changes in behavior that could partly explain these results (Van Winden, 2015; Passarelli and Tabellini, 2017). In the chapter I explore several mechanisms in these lines and provide evidence that indeed FCB performance affects euphoria and optimism, and this translates in more trust in the government, stronger group-identity and as a result preferences for more self-government.

Note that there are several differences between my setting and those examined in previous work which may contribute to our understanding of how identities and preferences are shaped around the world. First, most of the studies so far have looked at developing countries with high levels of diversity and inter-ethnic conflict. In my case, I study a consolidated democracy with hundreds of years of history and, while there has been increasing social unrest, there have been no recent episodes of conflict. Hence, my findings may be more easily extrapolated to other democracies with multiple "national" realities. Second, I examined a club that has been increasingly explicit about the political views it embraces, serving as an example of the consequences of politicizing sports.

Segregation and well-being

The country in which one is born is the most important factor determining an individuals' economic prospects. The gaps in income per capita between the richest and poorest countries in the world exceeds a ratio of a hundred. One could think that within the same

country, and after taking into one's account family background, individuals would have enjoy similar living standards, this view however could not be further away from reality. The neighborhood in which a child grows up is a major determinant of her future well-being. In the case of the US, Chetty et al. (2018) show that even conditioning on your parents' characteristics, there are stark differences in economic and life prospects depending on your childhood neighborhood. While a black boy from a low income family in Pocatello (Idaho) has 20% chances of being incarcerated during his lifetime, for a similar one from the neighboring county of Logan the chances are only 0.7%. These differences are explained by a combination of factors among which school peers and therefore school segregation play a prominent role.

The third chapter of this thesis aims at exploring the effects of school segregation on long-term well-being. To this respect, it exploits a major educational reform under Pinochet's dictatorship, which introduced school vouchers in 1981. As a result of the reform, in just 5 years the number of private schools doubled and the share of students enrolled in these establishments rose rapidly from 15% to over 35%. Since most affluent parents were more likely to send their children to private schools, this raise in private enrollment came hand in hand with a significant increase in school segregation. This piece of research presents several contributions such as focusing on the impact of school segregation on a broad measure such as subjective well-being, since the reform was nation-wide, it looks at the whole population and not just minorities, and finally, it examines its long-run effects decades after the reform took. Furthermore, it makes uses of administrative data of enrollment in the 1980s to appropriately measure the response to the policy change, and a research design comparing individuals born in the same place in different years to derive causal implications.

Most of the literature on school segregation has focused on their impact on educational attainment (Guryan, 2004; Johnston et al., 2007; Reber, 2010; Billings et al., 2014) and crime (Kling et al., 2005: Weiner et al., 2009; Billings et al., 2014). For instance, Billings et al. (2014) exploit how the abolition of a busing policy and strict rules that increase racial diversity in schools affected children's outcomes. They show that following the reform school segregation increased considerably because neighborhoods were highly segregated, and that this increase in school stratification had negative effects on boys test scores and graduation rates and raised crime among minority boys. In an exploratory work, Echenique et al. (2006) examine the relationship between school racial segregation and several outcomes during youth. Segregation in this context seemed to reduce the probability of inter-ethnic couples but did not have any significant effect on subjective well-being. Some limitations of this body of literature are that they just consider the case of the US, segregation is measured based only on ethnic origin, and they look solely at contemporaneous or young outcomes. Recent research has assessed the long-term impact of moving to richer neighborhoods on outcomes such as health, subjective well-being and earnings (Sanbonmatsu et al., 2012; Ludwig et al. 2012; Chetty et al., 2015). Despite relying on a randomized intervention, this research has produced inconclusive findings and lacks general equilibrium considerations.

There has been also numerous studies examining the impact of educational reforms and,

in particular, those related to school choice as a leading cause for school segregation. Some papers have found that flexible school choice tends to lead to increased segregation and exacerbate inequalities between children of different social background. Some theoretical models, such as Epple and Romano (1998) and MacLeod and Urquiola (2015), suggest that vouchers could lead to higher stratification (both in terms of family background and pupils' ability). Subsequent work has shown indeed that school vouchers tend to cause higher school segregation in a variety of contexts. There is evidence for the US (Brunner, Imazeki and Ross, 2010; Figlio, Hart and Metzger, 2010; Chakrabarti, 2013), Sweden (Böhlmark, Holmlund and Lindahl, 2016), Chile (Hsieh and Urquiola, 2006), and Kenya (Lucas and Mbiti, 2012), among others.

While past research has analyzed the effect of the Chilean voucher program on educational performance and stratification, little is known about other outcomes and the long-term consequences of the reform. The literature tends to find mixed results regarding educational quality and students' performance, but if any, the impact has been modest. The lack of appropriate data before and after the reform and a convincing research design to derive causal estimates has impeded reaching solid conclusions. By contrast, there seems to be a consensus pointing to a large increase in school stratification due to the voucher program (Hiseh and Urquiola, 2006; McEwan, Urquiola and Vegas, 2008; Elacqua, 2012; Valenzuela, Bellei and De los Rios, 2013; Santos and Elacqua, 2016).

The research presented in this third chapter makes several contributions. First, it is interested in school segregation rather than residential segregation, and on the basis of socio-economic status rather than race. These distinctions are important because in many settings, changes in school segregation come hand in hand with changes in residential segregation and it is difficult to disentangle the contribution of each separately. Similarly, while is often the case that minorities have lower earnings, this may not be necessarily true. In my analysis, I compare individuals who were born in the same municipality and thus were exposed to the same level of residential segregation, but depending on the year they enter school, they will be exposed to higher school segregation due to the 1981 voucher reform. While there are ethnic minorities in Chile, they represent a much lower share than in the US population.

Second, the chapter focuses on a developing country with extreme levels of educational inequalities and segregation. Chile is one of the countries in the world with a largest presence of the private sector in education. More than 50% of students are enrolled in private schools. Given that there is a strong correlation between family income and the type of establishments children attend, it does not come as surprise that the country ranks among the highest in terms of school segregation (OECD, 2014). Analyzing a developing country is relevant because these are precisely the ones lacking educational resources and who could be more interested in encouraging private schools via a voucher system alike Chile in the 1980s.

Third, it assesses the potential impact on an understudied and broad outcome: subjective well-being. The past decades have seen a growing amount of research relying on subjective

24

well-being data and its importance for public policy starts to be recognized (Dolan, Layard, Metcalfe, 2011; Frijters et al., 2020). Subjective well-being data is particularly useful when analyzing the consequences of public policies as it encompasses several dimensions of welfare into a single indicator. Moreover, it may unmask changes in well-being that are not perceived by using other economic data. In the context of studying segregation, it it becomes even more relevant given its potential effects on subtle aspects such as the network of friends and contacts, job insecurity and satisfaction, mental health, etc. Another crucial aspect is that to fully understand the consequences of educational reforms one needs to comprehensively analyze their potential long-term consequences. In this respect, I assess the impact of the 1981 reform three decades after it took place, implying that if individuals' well-being is still affected, the potential gains (cost) of educational reforms cannot be evaluated only on a short-term basis.

Finally, most of the research on school segregation focus on particular settings such as a city or a county experiencing a reforms. Since such policies can have all type of consequences on neighboring areas, to have a comprehensive view of its impacts they should be taken into account. In my case, given the nation-wide nature of Chilean reform, all individuals were affected and by analyzing data of the whole country I can take into account all the potential spillovers and general equilibrium effects generated by the increase in segregation. The latter are indeed crucial if one aims at understanding the potential consequences of scaling up similar policies.

Outline of the Thesis

This thesis exploits quasi-natural experiments, both historical and contemporary, to provide causal evidence on broad questions at the frontier of the development and political economy literature. In particular, it delves into the long-run effects of emigration in the communities of origin, the formation and drivers of national identities and political preferences, and the long-term impacts of school segregation on well-being.

One of the most important questions about migration is how it affects the communities of origin. In particular, economists have paid special attention to the consequences for the stock of human capital (Beine, Docquier and Rapoport, 2008). Now, establishing a causal relationship between migration and education in the sending communities is challenging because of the existence of factors affecting both migration decisions and investments in education and the possibility of reverse causality (Gibbons, McKenzie and Stillman, 2011). While researchers have implemented a variety of strategies to address these concerns, the approaches proposed often incorporate other caveats and problems of external validity. Additionally, a common feature of the past literature is that data limitations have made most studies to adopt a short-term view and analyze only the effects of migration in the very short run.

A long-term perspective is particularly relevant in the context of human capital accumulation because the impact of emigration can easily change over time, both in terms of size

and sign. Moreover, there are certain mechanisms that only play a role in the long run. As highlighted by the theoretical and empirical contributions (McKenzie and Rapoport, 2010; Dustman and Weiss,2007), the type of selection into emigration, the incentives to acquire education and the degree of return migration are all key factors to understand the effects of migration at origin, and they have an influence at different stages. Beyond that, migrants may also have a long-run impact on their communities of origin through more subtle mechanisms. For instance, they can contribute to the financing of public goods such as schools or act as drivers of cultural change and contribute to shape norms that encourage investments in human capital of the future generations.

The first chapter of this thesis revisits the question of how emigration affects education formation at origin with a historical and long-term approach. In particular, it examines the impact of mass emigration on human capital accumulation in the sending communities throughout an entire century. To do so, I focus on one the periods of largest migration flows in modern history: the Age of Mass Migration (Hatton & Williamson, 1998). Between 1850 and 1930, over 40 million Europeans crossed the Atlantic towards the Americas. In order to zoom in and be able to perform a causal analysis I take as a laboratory the Galician diaspora. Over the 1900-1930 period, more than 1.1 million people left this Spanish region towards Latin America, a figure representing 58% of its 1900 population. In relative terms, Galicia had higher emigration rates than those in any other European country at the time. The magnitude of this historical event, together with the similarities between Galicia in 1900-1930 and other developing countries today, make it an exceptional setting to study the short and long-run effects of mass emigration.

I combine newly digitized data from multiple historical sources (e.g., population censuses, embarkation lists, migrants' associations, among many others) with modern administrative and survey data to construct a unique database of Galician municipalities from 1860 until today. A main innovation is that I propose and validate a novel proxy for migration at the local level: missing residents. In particular, I rely on the share of men who were absent from their households the day of the census. I perform several checks to support that this is a good proxy for the stock of emigrants.

To perform a causal analysis I implement a research design in which I instrument the share of missing men by two plausibly exogenous sources of variation. To infer the intensity of migration in a given municipality, I use a proxy for pioneer emigration measured several decades before the mass emigration era. To infer the timing of migrant flows, I rely on a time-varying pull factor at the local level. More precisely, since most pioneer migrants were men I use the adult sex ratio measured in 1860 as proxy for this early flows. As a pull factor, I construct a decadal measure of GDP per capita growth at the average destination for each municipality using data on the share of individuals emigrating to each country. Using this empirical strategy, I analyze the drivers of emigration at the local level, its short and medium-run effects on literacy rates (1900-1930), and its long-run impact on different measures of educational attainment (1930s-2010s). Furthermore, I leverage other data sources to explore several mechanisms including the role played by migrants' associations at origin and the potential effect of migrants as a driver of change in values.

I find that emigration had a significant negative short-run effect on literacy rates at origin. This effect is present for both men and women and is explained by the fact that migrants were positively selected, i.e. (with higher literacy rates than the average of the population). Now, one decade later the impact of emigration becomes positive and significant meaning that it led to net gains in literacy rates. These positive effects are however only present for men. I discuss several channels that could explain these results. First, over 60% of migrants return to Galicia which would partly compensate the composition effect in the short-run. Secondly, emigration seem to have increase literacy rates among young children.

I then analyze the long-term consequences of the Galician diaspora for human capital at origin. Using data by cohorts from contemporary census I find that emigration led to higher rates of primary completion among individuals born in the 1930s, 1940s and later. I also find positive effects on secondary completion for certain cohorts and college completion in the most recent census. These findings highlight the importance of considering long horizons and several outcomes to have a clear picture of the effects of emigration at origin.

I also explore novel mechanisms to provide evidence on the long-term persistence of human capital. I first document that Galician funded hundreds of associations with an educational goal which financed the constructions of schools in their hometowns. Combining data on these schools and census data by cohorts I show that these investments raised educational achievement of the children exposed to them. Anecdotal evidence suggests that migrants and their associations were active factors of diffusion of information and norms about the importance of education. I use survey data in which individuals report their valuation of education and effort to shed light on this channels. I find that individuals in municipalities more exposed to emigration a century earlier today display higher valuation for education.

The findings of this chapter entail important implications for both research on migration and public policy. First, they highlight the relevance of taking a long-term view, as the impact of migration and the individuals affected may change over time in unexpected ways. For instance, while in the short-run both emigration depressed literacy of both men and women due to selection, in the following decades there was a positive impact on men's educational attainment only, and in the long run both men and women benefited similarly. They also stress the need to propose and evaluate other mechanisms beyond monetary remittances, such as collective remittances in the form of investments in public goods and social remittances. My results suggest that fostering migrants engagement with their communities of origin (e.g., through local associations, information campaigns, fundraising) can be a promising strategy to promote human capital accumulation and economic development at origin. Innovative public policies could be implemented to strengthen these links and encourage the transmission of collective remittances targeting specific goals. Finally, my findings shed new light on the impact of migration on the adoption of new cultural norms, a channel that may lead to persistent effects in the long run. Future work should examine in more depth the interplay between these different mechanisms, and explore the long-run impact of emigration on other social and economic aspects in the communities of origin.

The second chapter explores the connection between identity and political preferences,

and how they may respond to emotional shocks. There is a large literature highlighting how identity cleavages can be a major source of political and economic instability Besley and Reynal-Querol, 2014). In response to this, States have historically engaged in nation-building policies, fostering a common sense of identity through mass schooling and educational reforms (Bandiera et al., 2019; Cantoni et al., 2017; Clots-Figueras and Masella, 2013), propaganda and the use of media (Blouin and Mukand, 2019), and even mass re-allocations (Bazzi et al., 2019). Similarly, emotional shocks, especially those involving collective experiences, can also shape identity and values in complex ways (Dell and Querubin, 2018; Depetris-Chauvin et al., 2020).

In this respect, one of the most powerful forces driving individuals' passions and sentiments are sports. Football, in particular, has often been associated with the construction of ethnic and national identities, with countless examples of clubs fostering and embracing certain identities (Foer, 2004; Shobe, 2008a,b; Úbeda-Colomer et al., 2017) In a context of strong cleavages, this framework of rivalry, of opposition of positions, fueled by the emotional component of sports, can intensify differences between two entities in a context of strong cleavages (Ramonet, 1998), as exemplified by conflicts that had been fueled to football rivalries (Úbeda-Colomer et al., 2017).

This paper examines whether a close connection between a sports club and a feeling of identity can influence national identity, political preferences and lead to political change. I focus on the case of Catalonia and Football Club Barcelona (FCB), one of the most acclaimed clubs worldwide, and which has historically embraced and promoted Catalan identity. In fact, the hypothesis that FCB success could invigorated pro-secession sentiments and strengthened the Catalan independence process had already been proposed by some intellectuals as early as the 1990s (Vázquez-Montalbán, 2005).

The main data used comes from two opinion surveys conducted regularly in Spain by CEO (Centro d'Estudis d'Opinó) and CIS (Centro de Investigaciones Sociológicas). The CEO barometer surveys around 1500 people in Catalonia every quarter while the CIS barometer surveys around 2500 people in Spain every month. Both surveys include an identity question in which individuals have to describe themselves as "Only Spanish", "More Spanish than Catalan", "As Spanish as Catalan", "More Catalan than Spanish" or "Only Catalan". They also ask individuals about their preferences regarding State decentralization and the possibility of regions to become independent. I match the survey data with detailed information on FCB matches and trophies based on the date of the interviews. Given the differences between the CEO and CIS surveys, I use them to conduct two separate but complementary analyses: one focusing on the short-run impact of FCB matches and another looking at the short-run effect of FCB titles at the end of the season.

In order to arrive to the causal effects of FCB matches or FCB titles I exploit the exogenous timings of interviews around the events. In this sense, the fact that one individual is surveyed right before or after defeat is random. The same occurs with individuals surveying in a year in which FCB win more or less titles. I perform several analysis that support the assumption that *treated* and *control* individuals are comparable on observables.

Leveraging the survey data with information on FCB matches and trophies between 2009 and 2019, I find that FCB performance leads to sharp (temporary) changes in Catalan identification, preferences on decentralization and support for secession in Catalonia. Individuals surveyed one day after a defeat are 11% less likely to report a strong feeling of Catalan identity and are 15% less likely to support Catalan independence. By contrast, FCB titles at the end of the season increase identification and support for regional self-determination by around 11%. These effects persist for a couple of weeks before vanishing. My results are in line with the findings of Depetris-Chauvin et al. (2020) in the context of African football, but the implications are however the opposite: football can contribute to both unite and disunite a country. Exploiting data on FCB fans across municipalities, I find evidence that match defeats just before elections have a significant impact decreasing political participation.

I carry out several heterogeneity analysis and explore potential mechanisms for this effects. In particular, I look at proxies for changes in mood much as life satisfaction, optimism, trust on politicians, etc. I find that FCB titles lead to changes in optimism, improving individuals' economic and political prospects, and also lead to less mistrust in the government. These results support that idea that what is driving the changes in identity and preferences are short-run alterations in euphoria.

The findings of this chapter carry important policy implications. First, they demonstrate that individuals beliefs and preferences can be affected by random shocks through changes in mood. Furthermore, despite the temporary nature of the changes in identity and preferences caused by football, they could translate into permanent changes by affecting electoral results. All in all, the results call for the need to separate as much as possible sports from politics, and to be cautious when holding high-stake decisions as they may be influenced by emotional shocks.

The third chapter of this thesis examines the long-term effects of being exposed to school segregation during childhood. There has been a large number of studies exploring the consequences of school segregation for many dimensions such as educational attainment (Guryan, 2004; Johnston et al., 2007; Reber, 2010; Billings et al., 2014) and crime (Kling et al., 2005: Weiner et al., 2009; Billings et al., 2014). Yet, there is not empirical evidence on its potential long-run consequences on well-being.

To study the long-term effects of school segregation I take advantage of a unique historical event that shaped the Chilean educational system for decades. In 1981, the military government of Pinochet introduced a voucher system that allowed children to attend subsidized private schools without having to pay no fees. As a result, between 1981 and 1988 the number of private schools doubled and the share of children attending this type of establishments sore from 15 to 35 percent. Yet, this exodus was characterized by middle and upper-class students leading to a sharp increase in school segregation.

I use administrative of enrollment across municipalities before and after the reform to obtain a precise measure of the extent of privatization during the 1980s. Given the tight connection between parental income and the likelihood of children attending a subsided

private school, I interpret the raise in private enrollment between 1981 and 1988 as a proxy for the rise in segregation at the time. I then link the administrative data of enrollment with detailed individual survey data for the whole country for the years 2011 and 2013 based on individuals' birth municipality. Importantly, the survey data contains a standard subjective well-being question which is the focus of the analysis. Individuals are asked to rate their life satisfaction over the past weeks on a 1-10 scale.

To identify the causal effect of the 1981 of the reform I perform an analysis that exploits both variation across cohorts and space. In a nutshell, I compare individuals born in the same place but in different years, which determines their exposure to the reform. Individuals born in the early 1970s and thus attending school during the 1980s, while segregation was on the rise, are defined as the treated group. Individuals born in the early 1960s, and thus already out of compulsory education by 1981, make the control group or counterfactual. I then look how the gaps in life satisfaction between these two groups are influenced by the degree of school privatization experienced during the 1980s. I also use information on parental education to categorize individuals according to their family background and therefore their likelihood of having attended a private school.

I find evidence that the reform had long-lasting effects on life satisfaction and that it increased inequalities. Among individuals with more disadvantaged family backgrounds, a 10 percentage point increase in subsidized private enrollment when they were young is associated with a fall in adult life satisfaction in 5.9 percent of a standard deviation. This impact is equivalent to the effect of decreasing household income by around 30 percent or reducing education by 3 years. By contrast, those with most advantaged background seem to have benefited from the reform although the evidence is less robust. The impact differs markedly depending on sex and migration patterns. The negative effect on low family background individuals is entirely driven by those who stayed in their municipality of birth. The positive effect for high family background individuals is larger and significant among men and people who migrated internally to other municipalities. Focusing on those who moved, low family background individuals ending up in high privatization areas are substantially worse-off. In this case, a 10 percentage point increase in private enrollment is associated with a reduction in adult life satisfaction of 11.1 percent of a standard deviation.

I carry out a number of additional analysis to explore the robustness of the results. For instance, I alter the cohort windows considered as treated and control groups as if the reform had taken place in a different year and I also assess alternative measures of exposure taking into account the number of years individuals attended school after the reform. I also explore the potential effects on other outcomes such as schooling, earnings, and health, without finding any evidence that reform had a long-term impact on these dimensions.

The findings of this chapter illustrate how a policy meant to benefit the poorest students by providing them the opportunity to attend better schools had pernicious effects due to its bad design. Choice was never a true option for the most vulnerable families, and as result, they could not effectively benefit from the reform. The fact that this policy change had a permanent effect on adult life satisfaction calls for a well-thought design of educational policies, otherwise, even well-intention measures may compromise people's well-being for

the rest of their lives. It also challenges the view that educational policies should only be assessed by their impact on children's learning. Education years are critical for the formation of identities, beliefs and values, peers' networks, etc., aspects that may not be fully malleable later on. Future research should explore in depth the mechanisms underlying the negative effect on individuals from disadvantaged families to shed light on how such effects could have been avoided.

Bibliography

Acemoglu, D., S. Johnson, and J. A. Robinson. "The colonial origins of comparative development: An empirical investigation." *American Economic Review* 91, no. 5 (2001): 1369–1401.

Aggarwal, R., A. Demirgüç-Kunt, and M. S. M. Pería. "Do remittances promote financial development?" *Journal of Development Economics* 96, no. 2 (2011): 255–264.

Ahlerup, P., and O. Olsson. "The roots of ethnic diversity." *Journal of Economic Growth* 17, no. 2 (2012): 71–102.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf. "The economics of density: Evidence from the Berlin Wall." *Econometrica* 83, no. 6 (2015): 2127–2189.

Alesina, A., and E. L. Ferrara. "Ethnic diversity and economic performance." *Journal of economic literature* 43, no. 3 (2005): 762–800.

Alesina, A., and E. Zhuravskaya. "Segregation and the Quality of Government in a Cross Section of Countries." *American Economic Review* 101, no. 5 (2011): 1872–1911.

Algan, Y., C. Hémet, and D. D. Laitin. "The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation." *Journal of Political Economy* 124, no. 3 (2016): 696–733.

Angrist, J. D. "Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records." *The American Economic Review*, 1990, 313–336.

Angrist, J. D., and A. B. Krueger. "Instrumental variables and the search for identification: From supply and demand to natural experiments." *Journal of Economic perspectives* 15, no. 4 (2001): 69–85.

Angrist, J. D., and J. Pischke. "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics." *Journal of economic perspectives* 24, no. 2 (2010): 3–30.

Antman, F. M. "Gender, educational attainment, and the impact of parental migration on children left behind." *Journal of Population Economics* 25, no. 4 (2012): 1187–1214.

Bandiera, O., M. Mohnen, I. Rasul, and M. Viarengo. "Nation-building through compulsory schooling during the age of mass migration." *The Economic Journal* 129, no. 617 (2019): 62–109.

Banerjee, A., L. Iyer, and R. Somanathan. "History, social divisions, and public goods in rural India." *Journal of the European Economic Association* 3, nos. 2-3 (2005): 639–647.

Barsbai, T., H. Rapoport, A. Steinmayr, and C. Trebesch. "The effect of labor migration on the diffusion of democracy: evidence from a former Soviet Republic." *American Economic Journal: Applied Economics* 9, no. 3 (2017): 36–69.

Batista, C., and P. C. Vicente. "Do migrants improve governance at home? Evidence from a voting experiment." *The World Bank Economic Review* 25, no. 1 (2011): 77–104.

Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong. "Unity in diversity? how intergroup contact can foster nation building." *American Economic Review* 109, no. 11 (2019): 3978–4025.

Becker, S. O., and A. Ferrara. "Consequences of forced migration: A survey of recent findings." *Labour Economics* 59 (2019): 1–16.

Becker, S. O., I. Grosfeld, P. Grosjean, N. Voigtländer, and E. Zhuravskaya. "Forced migration and human capital: evidence from post-WWII population transfers." *American Economic Review* 110, no. 5 (2020): 1430–63.

Beine, M., F. Docquier, and H. Rapoport. "Brain drain and human capital formation in developing countries: winners and losers." *The Economic Journal* 118, no. 528 (2008): 631–652.

Beine, M., F. Docquier, and H. Rapoport. "Brain drain and economic growth: theory and evidence." *Journal of Development Economics* 64, no. 1 (2001): 275–289.

Besley, T., and M. Reynal-Querol. "The legacy of historical conflict: Evidence from Africa." *American Political Science Review*, 2014, 319–336.

Bhagwati, J. N., and M. Partington. *Taxing the Brain Drain: Theory and empirical analysis.* Vol. 2. North-Holland, 1976.

Billings, S. B., D. J. Deming, and J. Rockoff. "School segregation, educational attainment, and crime: Evidence from the end of busing in Charlotte-Mecklenburg." *The Quarterly Journal of Economics* 129, no. 1 (2014): 435–476.

Blouin, A., and S. W. Mukand. "Erasing ethnicity? Propaganda, nation building, and identity in Rwanda." *Journal of Political Economy* 127, no. 3 (2019): 1008–1062.

Böhlmark, A., H. Holmlund, and M. Lindahl. "Parental choice, neighbourhood segregation or cream skimming? An analysis of school segregation after a generalized choice reform." *Journal of Population Economics* 29, no. 4 (2016): 1155–1190.

Brunner, E. J., J. Imazeki, and S. L. Ross. "Universal vouchers and racial and ethnic segregation." *The Review of Economics and Statistics* 92, no. 4 (2010): 912–927.

Cantoni, D., Y. Chen, D. Y. Yang, N. Yuchtman, and Y. J. Zhang. "Curriculum and ideology." *Journal of Political Economy* 125, no. 2 (2017): 338–392.

Cantoni, D., and N. Yuchtman. *Historical Natural Experiments: Bridging Economics and Economic History.* Technical report. National Bureau of Economic Research, 2020.

Card, D., and G. B. Dahl. "Family violence and football: The effect of unexpected emotional cues on violent behavior." *The quarterly journal of economics* 126, no. 1 (2011): 103–143.

Carrington, M. W., and M. E. Detragiache. *How big is the brain drain?*, 98-102. International Monetary Fund, 1998.

Chakrabarti, R. "Do vouchers lead to sorting under random private school selection? Evidence from the Milwaukee voucher program." *Economics of Education Review* 34 (2013): 191–218.

Chen, W.-L., M.-J. Lin, and T.-T. Yang. *Curriculum and National Identity: Evidence from the 1997 Curriculum Reform in Taiwan.* Technical report. Working Paper, 2018.

Chetty, R., J. N. Friedman, N. Hendren, M. R. Jones, and S. R. Porter. *The opportunity atlas: Mapping the childhood roots of social mobility.* Technical report. National Bureau of Economic Research, 2018.

Chetty, R., and N. Hendren. "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects." *The Quarterly Journal of Economics* 133, no. 3 (2018): 1107–1162.

Clots-Figueras, I., and P. Masella. "Education, language and identity." *The Economic Journal* 123, no. 570 (2013): F332–F357.

Deaton, A., and N. Cartwright. "The limitations of randomised controlled trials." *VOX, CEPR policy portal.*, 2016.

Dell, M. "The persistent effects of Peru's mining mita." *Econometrica* 78, no. 6 (2010): 1863–1903.

Dell, M., and P. Querubin. "Nation building through foreign intervention: Evidence from discontinuities in military strategies." *The Quarterly Journal of Economics* 133, no. 2 (2018): 701–764.

Depetris-Chauvin, E., R. Durante, and F. Campante. "Building nations through shared experiences: Evidence from African football." *American Economic Review* 110, no. 5 (2020): 1572–1602.

Desmet, K., J. F. Gomes, and I. Ortuño-Ortín. "The geography of linguistic diversity and the provision of public goods." *Journal of Development Economics* 143 (2020): 102384.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg. "Culture, ethnicity, and diversity." *American Economic Review* 107, no. 9 (2017): 2479–2513.

Diamond, J. M., and D. Ordunio. *Guns, germs, and steel.* Books on Tape, 1999.

Dinkelman, T., and M. Mariotti. "The long-run effects of labor migration on human capital formation in communities of origin." *American Economic Journal: Applied Economics* 8, no. 4 (2016): 1–35.

Docquier, F., and A. Marfouk. "International migration by education attainment, 1990–2000." *International migration, remittances and the brain drain*, 2006, 151–199.

Docquier, F., and H. Rapoport. "Globalization, brain drain, and development." *Journal of Economic Literature* 50, no. 3 (2012): 681–730.

Dolan, P., R. Layard, and R. Metcalfe. "Measuring subjective well-being for public policy," 2011.

Droller, F. "Migration, population composition and long run economic development: Evidence from settlements in the pampas." *The Economic Journal* 128, no. 614 (2018): 2321–2352.

Duflo, E. "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment." *American Economic Review* 91, no. 4 (2001): 795–813.

Dustmann, C., and Y. Weiss. "Return migration: theory and empirical evidence from the UK." *British Journal of Industrial Relations* 45, no. 2 (2007): 236–256.

Echenique, F., R. G. Fryer Jr, and A. Kaufman. "Is school segregation good or bad?" *American Economic Review* 96, no. 2 (2006): 265–269.

Elacqua, G. "The impact of school choice and public policy on segregation: Evidence from Chile." *International Journal of Educational Development* 32, no. 3 (2012): 444–453.

Epple, D., and R. E. Romano. "Competition between private and public schools, vouchers, and peer-group effects." *American Economic Review*, 1998, 33–62.

Eren, O., and N. Mocan. "Emotional judges and unlucky juveniles." *American Economic Journal: Applied Economics* 10, no. 3 (2018): 171–205.

Feyrer, J., and B. Sacerdote. "Colonialism and modern income: islands as natural experiments." *The Review of Economics and Statistics* 91, no. 2 (2009): 245–262.

Figlio, D., C. M. Hart, and M. Metzger. "Who uses a means-tested scholarship, and what do they choose?" *Economics of Education Review* 29, no. 2 (2010): 301–317.

Foer, F. *How soccer explains the world: An unlikely theory of globalization.* HarperCollins New York, 2004.

Fouka, V. "Backlash: The unintended effects of language prohibition in US schools after World War I." *The Review of Economic Studies* 87, no. 1 (2020): 204–239.

Frijters, P., A. E. Clark, C. Krekel, and R. Layard. "A happy choice: wellbeing as the goal of government." *Behavioural Public Policy* 4, no. 2 (2020): 126–165.

Gennetian, L. A., L. Sanbonmatsu, L. F. Katz, J. R. Kling, M. Sciandra, J. Ludwig, G. J. Duncan, and R. C. Kessler. "The long-term effects of Moving to Opportunity on youth outcomes." *Cityscape*, 2012, 137–167.

Gibson, J., D. McKenzie, and S. Stillman. "The impacts of international migration on remaining household members: omnibus results from a migration lottery program." *Review of Economics and Statistics* 93, no. 4 (2011): 1297–1318.

Guiso, L., P. Sapienza, and L. Zingales. "Long-term persistence." *Journal of the European Economic Association* 14, no. 6 (2016): 1401–1436.

Harris, R., R. Johnston, and S. Burgess. "Neighborhoods, ethnicity and school choice: developing a statistical framework for geodemographic analysis." *Population Research and Policy Review* 26, nos. 5-6 (2007): 553–579.

Hatton, T. J., J. G. Williamson, et al. *The age of mass migration: Causes and economic impact.* Oxford University Press on Demand, 1998.

Hsieh, C.-T., and M. Urquiola. "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program." *Journal of public Economics* 90, nos. 8-9 (2006): 1477–1503.

Knudsen, A. S. B. *Those Who Stayed: Selection and Cultural Change during the Age of Mass Migration.* Technical report. Working Paper, 2019.

Lindo, J. M., I. D. Swensen, and G. R. Waddell. "Are big-time sports a threat to student achievement?" *American Economic Journal: Applied Economics* 4, no. 4 (2012): 254–74.

Lucas, A. M., and I. M. Mbiti. "Access, sorting, and achievement: The short-run effects of free primary education in Kenya." *American Economic Journal: Applied Economics* 4, no. 4 (2012): 226–53.

Ludwig, J., G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu. "Neighborhood effects on the long-term well-being of low-income adults." *Science* 337, no. 6101 (2012): 1505–1510.

MacLeod, W. B., and M. Urquiola. "Reputation and school competition." *American Economic Review* 105, no. 11 (2015): 3471–88.

Mankiw, N. G., D. Romer, and D. N. Weil. "A contribution to the empirics of economic growth." *The quarterly journal of economics* 107, no. 2 (1992): 407–437.

Margo, R. A. "The Economic History of Economic History: The Evolution of a Field in Economics," 2019.

Mayr, E. "The objects of selection." *Proceedings of the National Academy of Sciences* 94, no. 6 (1997): 2091–2094.

McAuliffe, M., B. Khadria, M. N. Céline Bauloz, et al. "World migration report 2020." *International Organisation for Migration*, 2020.

McEwan, P. J., M. Urquiola, E. Vegas, R. Fernandes, and F. A. Gallego. "School choice, stratification, and information on school performance: Lessons from chile [with comments]." *Economia* 8, no. 2 (2008): 1–42.

McKenzie, D., and H. Rapoport. *Can migration reduce educational attainment? Evidence from Mexico.* The World Bank, 2006.

———. "Self-selection patterns in Mexico-US migration: the role of migration networks." *the Review of Economics and Statistics* 92, no. 4 (2010): 811–821.

Michalopoulos, S. "The origins of ethnolinguistic diversity." *American Economic Review* 102, no. 4 (2012): 1508–39.

Michalopoulos, S., and E. Papaioannou. "The long-run effects of the scramble for Africa." *American Economic Review* 106, no. 7 (2016): 1802–48.

Miguel, E., and M. Kremer. "Worms: identifying impacts on education and health in the presence of treatment externalities." *Econometrica* 72, no. 1 (2004): 159–217.

Montalvo, J. G., and M. Reynal-Querol. "Ethnic polarization, potential conflict, and civil wars." *American Economic Review* 95, no. 3 (2005): 796–816.

Munyo, I., and M. A. Rossi. "Frustration, euphoria, and violent crime." *Journal of Economic Behavior & Organization* 89 (2013): 136–142.

North, D. C. "Institutions." *Journal of economic perspectives* 5, no. 1 (1991): 97–112.

Nunn, N. "Rethinking economic development." *Canadian Journal of Economics/Revue canadienne d'économique* 52, no. 4 (2019): 1349–1373.

———. "The long-term effects of Africa's slave trades." *The Quarterly Journal of Economics* 123, no. 1 (2008): 139–176.

Nunn, N., and L. Wantchekon. "The slave trade and the origins of mistrust in Africa." *American Economic Review* 101, no. 7 (2011): 3221–52.

Ozden, C., M. Wagner, and M. Packard. "Moving for Prosperity: Global Migration and Labor Markets." *Policy Research Report Overview, World Bank Group,(15)*, 2018.

Pascali, L. "The wind of change: Maritime technology, trade, and economic development." *American Economic Review* 107, no. 9 (2017): 2821–54.

Passarelli, F., and G. Tabellini. "Emotions and political unrest." *Journal of Political Economy* 125, no. 3 (2017): 903–946.

Reber, S. J. "School desegregation and educational attainment for blacks." *Journal of Human resources* 45, no. 4 (2010): 893–914.

Ricardo, D. *On the Principles of Political Economy and Taxation (1821).* Kessinger Publishing, 2009.

Rocha, R., C. Ferraz, and R. R. Soares. "Human capital persistence and development." *American Economic Journal: Applied Economics* 9, no. 4 (2017): 105–36.

Sachs, J. D., and A. M. Warner. "The curse of natural resources." *European economic review* 45, nos. 4-6 (2001): 827–838.

Santos, H., and G. Elacqua. "Segregación socioeconómica escolar en Chile: elección de la escuela por los padres y un análisis contrafactual teórico," 2016.

Sequeira, S., N. Nunn, and N. Qian. "Immigrants and the Making of America." *The Review of Economic Studies* 87, no. 1 (2020): 382–419.

Shobe, H. "Football and the politics of place: Football Club Barcelona and Catalonia, 1975–2005." *Journal of cultural geography* 25, no. 1 (2008): 87–105.

———. "Place, identity and football: Catalonia, catalanisme and football club Barcelona, 1899–1975." *National identities* 10, no. 3 (2008): 329–343.

Shrestha, S. A. "No man left behind: Effects of emigration prospects on educational and labour outcomes of non-migrants." *The Economic Journal* 127, no. 600 (2017): 495–521.

Smith, A. *The Wealth of Nations: An inquiry into the nature and causes of the Wealth of Nations.* Harriman House Limited, 2010.

Solow, R. M. "Technical change and the aggregate production function." *The review of Economics and Statistics*, 1957, 312–320.

Tabellini, G. "Culture and institutions: economic development in the regions of Europe." *Journal of the European Economic association* 8, no. 4 (2010): 677–716.

Testa, P. "The economic legacy of expulsion: Lessons from postwar Czechoslovakia." *Available at SSRN 3251541*, 2019.

Úbeda-Colomer, J., P. Molina, and M. Villamón. "La construcción de identidades colectivas a través del fútbol: cuando los nacionalismos juegan en los estadios= Collective identities construction through football: when nationalisms play in stadiums." *Materiales para la Historia del Deporte*, no. 15 (2017): 45–56.

Valencia Caicedo, F. "The mission: Human capital transmission, economic persistence, and culture in South America." *The Quarterly Journal of Economics* 134, no. 1 (2019): 507–556.

Valenzuela, J. P., C. Bellei, and D. d. l. Ríos. "Socioeconomic school segregation in a market-oriented educational system. The case of Chile." *Journal of education Policy* 29, no. 2 (2014): 217–241.

Vázquez Montalbán, M. "Fútbol. Una religión en busca de un Dios." *Barcelona, Debate*, 2005.

Voigtländer, N., and H.-J. Voth. "Nazi indoctrination and anti-Semitic beliefs in Germany." *Proceedings of the National Academy of Sciences* 112, no. 26 (2015): 7931–7936.

Weiner, D. A., B. F. Lutz, and J. Ludwig. *The effects of school desegregation on crime.* Technical report. National Bureau of Economic Research, 2009.

Introduction

Les grandes questions et l'histoire comme laboratoire

"Beaucoup de progrès dans les sciences sociales sont dus au génie de ceux qui ont dé-couvert, évalué et comparé de manière critique [...] des expériences naturelles dans des domaines où une expérience de laboratoire n'est pas pratique, voire impossible." – Ernst Mayr (1997, p.29)

Les recherches les plus remarquables en économie au cours des dernières décennies ont été caractérisées par une combinaison de réponses à des questions pertinentes avec les méth-odes les plus convaincantes. Cela a impliqué de se concentrer sur les grandes questions, d'exploiter des techniques ou des modèles de recherche qui fournissent des contrefactuels appropriés, et d'avoir une compréhension approfondie des contextes étudiés.

L'un des aspects qui a reçu le plus d'attention a tourné autour des sources du développe-ment économique. Pourquoi certains pays sont-ils plus riches que d'autres ? Que peut-on faire pour promouvoir la croissance économique et la prospérité ? Selon les mots de Robert Lucas (1998): *"[...] Les conséquences pour le bien-être humain qu'impliquent de telles questions sont tout simplement stupéfiantes : une fois que l'on commence à y penser, il est difficile de penser à autre chose."* Ce type de questions remonte aux économistes classiques tels qu'Adam Smith (1776) et David Ricardo (1817) et a reçu des contributions majeures dans la seconde moitié du XXe siècle (Solow, 1957 ; Mankiw, Romer et Weil, 1991). Malgré les progrès remarquables de nombreux pays en développement dans plusieurs domaines, tels que les investissements dans le capital physique ou l'accès à la santé et à l'éducation (Deaton, 2013), nous constatons encore aujourd'hui de grandes différences dans les niveaux de revenus à travers le monde, ce qui a suscité un vif débat sur les causes fondamentales du développement. Pour expliquer ces inégalités persistantes, certains économistes ont souligné le rôle joué par des facteurs tels que la géographie (Sachs et Warner, 1995 ; Dia-mond, 1998), les institutions (North, 1990 ; Acemoglu, Robinson et Johnson, 2001), et plus récemment la culture et son interaction avec la première (Tabellini ; 2010, Nunn ; 2011).

Ces derniers temps, non seulement le champ de la recherche économique s'est élargi, mais elle a également connu un changement rapide dans son approche, mettant davantage l'accent sur l'estimation crédible des effets de causalité. Aujourd'hui plus que jamais, les économistes s'efforcent d'évaluer soigneusement les données qu'ils exploitent, de constru-ire des contrefactuels significatifs et d'énoncer clairement les hypothèses qui sous-tendent

leur conception de la recherche, ainsi que de fournir des preuves qui peuvent étayer ces hypothèses. Cette "révolution de la crédibilité" (Angrist et Pishke, 2010), qui a commencé dans l'économie du travail et l'économie publique au début des années 1990 (Angrist, 1990, Angrist et Krueger, 1991), s'est rapidement étendue à des domaines très importants de l'économie du développement (voir les travaux de Duflo, 2001 ; Miguel et Kremer, 2004) et, plus récemment, à des domaines traditionnellement moins quantitatifs tels que l'histoire économique (quelques exemples importants sont Nunn, 2008 ; Dell, 2010).

Cette révolution de la crédibilité s'est accompagnée d'une augmentation extraordinaire de l'utilisation d'expériences naturelles historiques au sein de l'économie (Margo, 2020). La raison en est que l'histoire fournit un laboratoire idéal avec des sources de variation uniques qui permettent d'aborder les grandes questions de manière désinvolte (Cantoni et Yuchtman, 2020). En fait, les événements historiques permettent souvent de vérifier des hypothèses ou d'explorer des questions qu'il aurait été impossible d'aborder à l'époque moderne ou avec des données contemporaines. Un autre avantage de l'adoption d'une perspective historique et à long terme est qu'elle permet de laisser passer suffisamment de temps et donc d'examiner comment différentes dynamiques évoluent ou de prendre en compte les effets d'équilibre général.

Il y a eu une explosion d'études utilisant des modèles de recherche innovants et une variété de techniques tout en tirant parti des données historiques. Par exemple, des variables instrumentales originales telles que la direction des vents et leur importance pour différentes technologies de navigation, comme celles utilisées par Feyrer et Sacerdote (2009) pour étudier les impacts du colonialisme ou par Pascali (2017) pour explorer l'impact du commerce sur le développement. La distance à une frontière dans un modèle de discontinuité de régression pour comprendre les effets à long terme des institutions extractives (Dell, 2010) ou des économies d'agglomération (Ahlfeldt et al., 2015). Egalement des instruments de partage de la production qui tirent parti de la variation des facteurs de changement spatial et temporel (Sequeira, Nunn et Quian, 2020).

Au milieu du récent boom des évaluations d'impact et des essais contrôlés randomisés, certains économistes du développement élèvent la voix pour remettre en question l'utilité de ces approches sans réfléchir au contexte des études. Tout d'abord, il est souvent difficile de savoir si les particularités du contexte peuvent rendre les résultats difficiles à extrapoler à d'autres contextes, ce que l'on appelle communément la "validité externe" (Deaton et Cartwright, 2016). Deuxièmement, même sans s'interroger sur la validité externe, une mauvaise compréhension du contexte limite la capacité du chercheur à interpréter de manière exhaustive les résultats et leurs implications (Nunn, 2019). Comme le dit Nunn (2019), sans se rendre compte de l'importance du contexte local, les chercheurs auront une vision incomplète ou incorrecte des questions de recherche elles-mêmes, et une mauvaise compréhension des raisons pour lesquelles certaines politiques peuvent réussir ou échouer. Pour y remédier, il est nécessaire d'acquérir des connaissances plus spécifiques au contexte, d'effectuer un travail qualitatif et de collaborer avec des universitaires locaux dans tous les domaines.

Cette thèse ne fait pas exception à la règle puisqu'elle tente d'aborder les grandes questions

de la littérature sur le développement et l'économie politique. Plutôt que d'explorer un seul sujet et diverses questions qui l'entourent, chaque chapitre explore une question différente dans ce qui pourrait être considéré comme un document de recherche à part entière (d'où l'utilisation du mot "papier" dans les chapitres). De même, bien que l'approche soit toujours empirique, les techniques utilisées dans chaque chapitre sont également différentes, et particulièrement adaptées à la question et au contexte étudiés. Les chapitres suivants aborderont des aspects tels que la manière dont l'émigration affecte les communautés d'origine, les chocs émotionnels qui peuvent affecter le sentiment d'identité nationale et influencer les préférences politiques, ou encore les effets de la ségrégation scolaire sur le bien-être à long terme.

Comme exposé précédemment, cette thèse suit l'approche consistant à utiliser avec soin les données et les techniques économétriques afin d'établir des relations causales. Dans les chapitres, je mets en œuvre un large éventail de techniques économétriques telles que les instruments de partage de poste, voire les études et l'analyse de cohorte, entre autres.

Les chapitres 1 et 3 tirent parti des expériences historiques naturelles (à savoir la diaspora galicienne pendant les années 1900-1930 au chapitre 1 et une réforme majeure de l'éducation pendant la dictature du Chili au chapitre 3) tandis que le chapitre 2 repose sur un cadre quasi expérimental dans lequel les individus sont exposés de manière aléatoire à des chocs inattendus (par exemple, la perte d'un match du Club de football de Barcelone, un porte-drapeau de l'identité catalane). Les chapitres 1 et 3 s'appuient tous deux sur des données historiques uniques recueillies pour les besoins de cette thèse. Dans le cadre d'une tâche héroïque de collecte de données, j'ai effectué plusieurs voyages en Espagne, notamment à Madrid et à Santiago, où j'ai recueilli des données historiques à partir des statistiques espagnoles sur les migrations vers l'Amérique latine, des recensements de l'éducation et de la population et de nombreuses autres sources. Au cours de ces visites, j'ai non seulement pris des milliers de photos de ces documents, mais j'ai également profité de l'occasion pour visiter certaines bibliothèques et lire différents livres, thèses et manuscrits, qui auraient été inaccessibles autrement. Dans le même esprit, juste avant d'entamer mon doctorat, j'ai passé plusieurs mois au Chili à la recherche de certaines données que la plupart des chercheurs considèrent comme "perdues à jamais". Après avoir pris contact avec des dizaines de personnes, j'ai finalement réussi à trouver ces documents "perdus" dans des archives du ministère de l'éducation dans la banlieue de Santiago. J'y ai trouvé des données sur les inscriptions scolaires pour toutes les municipalités chiliennes au cours des années 80, ce qui constitue la base de l'analyse de la réforme des bons d'études de 1981. Dans les deux cas, j'ai dû passer plusieurs mois à taper les données à la main dans des tessons de papier afin de constituer les ensembles de données qui m'ont ensuite permis d'effectuer les analyses.

Enfin, il est essentiel de souligner que cette thèse ne traite pas seulement de questions du type "qu'est-ce qui cause quoi", mais vise également à quantifier ces relations et à approfondir le "comment et pourquoi", en étudiant les mécanismes qui sous-tendent les conclusions principales. Pour guider cette analyse, elle s'appuie sur une compréhension approfondie du contexte de chaque étude et sur une lecture approfondie de la littérature dans d'autres domaines tels que l'histoire et la sociologie. Cette combinaison de questions de recherche ambitieuses, de méthodes rigoureuses et d'une approche multidisciplinaire sont

les traits distinctifs de cette thèse.

Littérature et principales contributions

Migration et capital humain

Une question fondamentale dans la littérature sur le développement est de savoir comment l'émigration affecte les pays d'origine. L'émigration favorise-t-elle ou entrave-t-elle le développement économique (Aggarwal, Demirgüç-Kunt, Pería, 2011) ? Appauvrit-elle ou augmente-t-elle le stock de capital humain (Beine, Docquier et Rapoport, 2008) ? Peut-elle contribuer à changer les institutions et à consolider les démocraties (Barsbai et al., 2017) ? Avec un nombre croissant de migrants au cours des dernières décennies et des perspectives d'augmentation des flux dans les années à venir (Nations unies, 2020), ces aspects sont aujourd'hui plus pertinents que jamais.

Un grand nombre de travaux ont porté sur l'examen des tendances et des modèles de migration internationale (Docquier et Marfouk, 2006). Les chercheurs se sont rapidement rendu compte que les migrants ont tendance à être sélectionnés de manière positive en termes d'éducation (c'est-à-dire avec un niveau d'éducation supérieur à la moyenne de la population) et que, par conséquent, leur départ pourrait réduire le stock de capital humain à l'origine (Carrington et Detragiache, 1998). Face à cela, certains économistes ont suggéré que les pays en développement mettent en place une "taxe sur les cerveaux" pour compenser le départ de ces personnes instruites (Bhagwati, 1976). Il s'avère que la sélection positive des migrants est plus la norme que l'exception. En fait, plus de 80% de tous les pays semblent afficher un tel schéma aujourd'hui (Banque mondiale, 2018) et c'est un phénomène qui s'est accentué au fil du temps (Docquier et Rapoport, 2012). En outre, les pays les plus pauvres ont tendance à présenter des taux d'émigration plus élevés et un degré plus élevé de sélection positive (Banque mondiale, 2018). La combinaison de ces trois faits semble créer une tempête parfaite, indiquant que l'émigration pourrait effectivement être préjudiciable au stock de capital humain dans les pays d'origine, en particulier pour les plus pauvres.

Pourtant, certains travaux réalisés il y a une vingtaine d'années ont déjà souligné que la relation entre l'émigration et l'éducation est beaucoup plus complexe qu'un simple effet de composition. Au-delà de l'effet direct causé par le départ des personnes ayant un certain niveau d'éducation, les perspectives d'émigration et les envois de fonds peuvent également influencer les investissements dans l'éducation. En effet, si les perspectives d'émigration constituent une incitation suffisamment forte à investir dans le capital humain mais que seul un nombre limité de ces personnes finissent par partir, l'impact net de l'émigration sur le stock de capital humain ne devrait pas être négatif, et pourrait même être positif (Beine, Docquier et Rapoport, 2001).

Ces dernières années, un grand nombre d'études ont été réalisées pour examiner les effets de l'émigration sur la famille restée au pays, et en particulier sur les résultats scolaires des enfants. Par exemple, Antmann (2012) a étudié la migration mexico-américaine et suggère

que l'émigration pourrait être bénéfique pour les enfants laissés derrière, en particulier pour les filles. Elle exploite les différences d'exposition à l'émigration en fonction de l'âge des enfants lorsque le parent migre. Les travaux de McKenzie & Rapoport (2006, 2010) dans le même contexte remettent en cause cette conclusion en indiquant que l'impact de l'émigration peut être plus complexe, et même dépendre des réseaux de migrants. En fait, si ces réseaux permettent aux personnes peu qualifiées de gagner des salaires plus élevés à destination, cela pourrait inciter les enfants à quitter l'école prématurément et avoir ainsi un impact négatif sur le stock de capital humain.

Ces études et d'autres du même type ont relevé le défi que les décisions en matière de migration ne sont pas exogènes. Tout d'abord, les ménages avec ou sans membre migrant ont tendance à être différents sur un grand nombre de points. Deuxièmement, le moment auquel un membre de la famille décide d'émigrer peut être corrélé avec d'autres chocs et conditions affectant le ménage. Enfin, dans certains cas, des ménages entiers émigrent, ce qui rend plus difficile l'obtention de données à leur sujet. Bien que les auteurs aient essayé de mettre en œuvre des modèles de recherche qui fournissent un meilleur contrefactuel, en d'autres termes, ce qui se serait passé si cette famille n'avait pas émigré, ou s'il n'y avait pas eu d'émigration de cette communauté, il reste des préoccupations quant à l'existence de facteurs de confusion affectant les flux migratoires. Pour surmonter ces limites, d'autres travaux se sont appuyés sur de riches enquêtes ou des cadres quasi-expérimentaux dans lesquels l'option d'émigrer est aléatoire. Par exemple, Batista, Lacuesta et Vicente (2011) tirent parti d'une riche enquête sur les ménages menée au Cap-Vert, qui contient les récits de migration de tous les membres du ménage, et estiment que les perspectives d'émigration augmentent le niveau d'éducation des individus. Gibbons, McKenzie et Stillman (2011) exploitent un système de loterie pour émigrer des Tonga vers la Nouvelle-Zélande afin d'examiner l'impact de l'émigration sur la famille restée au pays sur divers résultats, dont l'éducation. Ils ne constatent aucun impact significatif sur l'alphabétisation ou les années d'éducation, mais une mise en garde importante est que dans leur contexte, les migrants ne sont généralement pas les parents des enfants pour lesquels ils disposent de données, mais plutôt leurs oncles et tantes. Shrestha (2017) examine l'impact de l'introduction d'une exigence en matière d'éducation dans la sélection des Népalais pour rejoindre l'armée britannique et constate des effets positifs importants, qu'ils aient ou non été retenus dans leur candidature. Il documente ainsi que l'établissement d'une telle exigence a conduit à des gains nets en capital humain au sein des communautés concernées. Bien que ces études fournissent d'importantes informations, les particularités des milieux étudiés et leurs modèles de recherche posent de sérieuses questions sur la validité externe des résultats. En outre, elles sont généralement agnostiques quant aux mécanismes qui pourraient être à l'origine de leurs effets.

Une autre limite importante de la littérature précédente est qu'elle a eu tendance à se concentrer sur l'impact de la migration sur les migrants potentiels ou les enfants laissés derrière au moment où la migration a lieu. Cette perspective à court terme n'est pas idéale pour un certain nombre de raisons. Tout d'abord, les enfants de migrants peuvent émigrer eux-mêmes, de sorte que sans tenir compte de leurs futures trajectoires migratoires, il est impossible d'évaluer l'effet net sur le stock de capital humain. En outre, une part considérable des migrants peut retourner dans leur pays d'origine et, par conséquent, l'ampleur

de cet afflux et le degré de sélection dans la migration de retour peuvent jouer un rôle majeur. Il est essentiel de savoir quand les migrants ont tendance à revenir pour déterminer l'horizon temporel à analyser. Une étude qui a adopté une perspective à plus long terme est celle de Dinkelman et Mariotti (2016), dont je parlerai plus loin.

Les mécanismes dont il est question impliquent des changements dans la composition de la population (c'est-à-dire qui part, qui reste et qui revient), il existe cependant plusieurs autres mécanismes qui pourraient faire varier l'impact de l'émigration à court et à long terme. Par exemple, les migrants peuvent investir dans les biens publics de leur communauté d'origine, tels que des écoles ou des hôpitaux (Chauvet et al., 2014). Ces investissements peuvent à leur tour avoir un effet positif sur les résultats scolaires des enfants, et surtout, sur la communauté dans son ensemble et pas seulement sur les enfants de migrants. Cependant, ces investissements ne peuvent avoir lieu que lorsque les migrants ont accumulé suffisamment de richesses ou si un système de transferts collectifs est mis en place, par exemple par le biais d'associations de migrants dans leur ville d'origine. De même, les migrants peuvent s'engager dans des transferts sociaux liés à la valeur ou au rendement de l'éducation. En diffusant de telles normes ou informations, ils pourraient influencer les décisions des individus quant à la quantité de capital humain à acquérir ou à investir pour leurs enfants. Pourtant, ces processus de changement culturel peuvent prendre plusieurs générations pour se concrétiser. Il est important de noter qu'à long terme, l'émigration peut impliquer toutes sortes d'effets d'équilibre général qui ne peuvent être saisis que si un temps suffisant s'est écoulé et si l'on considère les communautés dans leur ensemble.

Le premier chapitre de cette thèse réexamine la question de l'impact de l'émigration sur l'accumulation de capital humain à l'origine, mais il introduit plusieurs innovations majeures par rapport à la littérature précédente. L'une des principales contributions est qu'elle adopte une perspective à très long terme et une approche historique. Pour ce faire, je me concentre sur l'un des plus grands épisodes migratoires de l'histoire moderne, l'Age of Mass Migration (1850- 1930), une période au cours de laquelle plus de 40 millions d'Européens ont traversé l'Atlantique en direction des Amériques (Hatton et Williamson, 1998). Je me penche en particulier sur l'émigration massive de Galiciens vers l'Amérique latine qui a eu lieu entre 1900 et 1930 et sur ses effets sur l'accumulation de capital humain pendant le reste du siècle.

En exploitant de nouvelles données historiques avec des données contemporaines couvrant plus de cent ans, je documente que l'émigration a conduit à des gains persistants de capital humain dans les communautés d'origine sur plusieurs générations. Plus important encore, je mets en œuvre un plan de recherche qui exploite les fluctuations des conditions économiques à destination dans une approche de variables instrumentales afin d'obtenir une variation exogène du moment de l'émigration dans les municipalités galiciennes et une approximation des réseaux de migrants passés pour déduire l'intensité de l'émigration. Avec cette stratégie et en prenant les municipalités comme unité d'analyse, je contourne les problèmes de la recherche passée qui devait traiter de la sélection des individus ou des ménages dans l'émigration.

Cette étude contribue ainsi à la recherche croissante qui explore les effets à long terme

de la migration. Des travaux récents ont documenté les effets positifs à long terme sur le développement économique et l'accumulation de capital humain dans les régions qui accueillent des migrants (Rocha, Ferraz et Soares, 2017 ; Droller, 2018 ; Sequeira, Nunn et Quian, 2020). En ce qui concerne les impacts à l'origine, la littérature a constaté des effets positifs sur l'innovation (Anderson, Karadja et Prawitz, 2020), et des résultats mitigés en termes de développement économique et de formation de capital humain (Dinkelman et Mariotti ,2016 ; Testa, 2020). J'apporte de nouvelles preuves que, malgré une perte à court terme, l'émigration massive de Galice a entraîné des gains en capital humain qui ont persisté pendant plus de cent ans. Pour comprendre la variété des résultats, il est essentiel de comprendre le contexte de chaque étude. Dinkelman et Mariotti (2016), par exemple, examinent la migration circulaire de la Tanzanie vers les mines sud-africaines, qui, pour des raisons politiques, a soudainement cessé de forcer tous les migrants à rentrer chez eux. Étant donné que tous les migrants ont dû rentrer et qu'ils ont reçu tous leurs revenus à leur arrivée, ce contexte est particulièrement propice à un impact positif sur la formation du capital humain. Dans le cas de Testa (2020), il étudie un épisode de migration forcée (Allemands expulsés de Tchécoslovaquie après la Seconde Guerre mondiale), c'est-à-dire qu'un groupe particulier de la population a été expulsé de ses terres et n'a pas été autorisé à y revenir. Il constate que cet épisode a eu un effet négatif à long terme sur le capital humain dans les régions qui ont connu un exode plus important.

À cet égard, la diaspora galicienne offre un meilleur contexte pour analyser les effets de l'émigration à l'origine. Premièrement, au début du XXe siècle, la Galice partageait de nombreuses caractéristiques avec d'autres pays en développement aujourd'hui, telles qu'une économie reposant presque exclusivement sur l'agriculture et un très faible niveau de capital humain et d'infrastructures éducatives. En outre, elle était similaire à d'autres flux migratoires en provenance d'Europe du Sud à l'époque des migrations de masse et à d'autres flux migratoires aujourd'hui, caractérisés par des migrants masculins qui quittent souvent leur famille et maintiennent ainsi des liens étroits avec leur communauté d'origine, ce qui contraste nettement avec les épisodes de migration forcée (Becker et Ferrara, 2019). Enfin, les décisions d'émigrer, de revenir ou d'envoyer des fonds n'ont pas été influencées par des facteurs institutionnels ou des guerres et ressemblent donc à celles de la plupart des épisodes migratoires actuels.

Une dernière contribution de ce travail concerne les mécanismes explorés. J'examine deux nouveaux canaux par lesquels l'émigration pourrait affecter le capital humain à long terme, à savoir les investissements dans les biens publics et les changements culturels. Inspiré par les travaux examinant le rôle des institutions non gouvernementales finançant les biens publics et contribuant au développement local (Valencia-Caicedo, 2019), ce chapitre étudie la contribution des associations de migrants. Je documente le fait que les migrants galiciens ont créé des centaines d'associations dans le but d'investir dans la construction d'écoles dans leurs communautés d'origine, ce qui pourrait être considéré comme une forme pionnière d'aide au développement. J'analyse les facteurs qui influencent l'origine de ces associations, leur décision de construire des écoles dans leurs villages, et leur impact sur le niveau d'éducation et la poursuite de la migration.

Ce travail se rapporte également à une littérature récente sur la migration et la culture. Par

exemple, Knudsen (2019) montre que les migrants des pays scandinaves étaient en moyenne plus individualistes et que, par conséquent, leur départ a élevé le niveau de collectivisme à l'origine. Becker et ses collaborateurs (2020) montrent que les migrants forcés polonais sont devenus par la suite moins matérialistes et ont investi davantage dans l'éducation, étant donné que celle-ci est de nature transférable. Je contribue à cette littérature en examinant comment l'émigration a affecté les croyances sur la valeur de l'éducation et de l'effort. Plutôt que le résultat de la sélection des migrants ou de la migration forcée (Becker et al., 2020), mes conclusions semblent suggérer qu'une exposition à un environnement où l'éducation est hautement valorisée peut amener les migrants à changer leurs normes et à les transmettre à leurs communautés d'origine.

Identité national et préférences politiques

La plupart des pays du monde, en particulier les pays en développement, présentent un niveau élevé de diversité et de forts clivages identitaires. Ces clivages peuvent également prendre la forme de différences de langue, de religion, etc. Il est très important de comprendre la nature de ces clivages, car ils peuvent conduire à des conflits et entraver la croissance économique (Alesina et La Ferrara, 2005). De même, ils peuvent engendrer des méfiances entre les individus de différents groupes, des sociétés moins cohésives et même un sous-investissement dans les biens publics (Guiso et al., 2016). Étant donné les vastes implications des clivages ethniques, il n'est pas surprenant que les États modernes se soient engagés dans diverses politiques de construction de la nation afin de favoriser un sentiment commun d'identité. Parmi celles-ci, on peut citer la propagande (Blouin et Mukand, 2019), la scolarisation de masse (Bandiera et al., 2013), et même le mélange des populations (Bazzi et al., 2019). Pourtant, même s'il réussit, l'héritage des clivages ethniques peut persister pendant des générations (Besley et Reynal-Querol, 2014).

Une grande partie de la littérature a exploré les conséquences économiques de la diversité et de la ségrégation ethniques, qui sont au cœur des clivages identitaires. Par exemple, Montalvo et Reynal-Querol (2005a, 2005b) montrent empiriquement que la polarisation ethnique peut réduire le développement économique par son effet sur l'apparition de conflits, en réduisant les investissements et en affectant les dépenses gouvernementales. Un certain nombre d'études montrent qu'une plus grande diversité ethnique tend à entraîner une baisse des investissements dans les biens publics (Banerjee, Lakshimi et Somanathan, 2005 ; Algan, Hémet et Laitin, 2016 ; Desmet, Gomes et Ortuño-Ortín, 2020). Cela s'explique par le fait que les sociétés où la ségrégation ethnique est plus importante ont tendance à afficher des niveaux de confiance interethnique plus faibles (Alesina et Zhuravskaya, 2011) et plus de difficultés à coopérer (Guiso et al., 2016). Des recherches récentes ont souligné que les identités et les valeurs culturelles interagissent de manière complexe, et que la manière dont ces deux aspects se chevauchent peut avoir des implications économiques et politiques différentes (Desmet, Ortuño-Ortín et Wacziarg, 2017). D'où proviennent les clivages identitaires ? Les niveaux élevés de diversité et de polarisation ethniques ont souvent été le résultat de processus historiques impliquant l'installation de populations et les flux migratoires (Ahlerup et Olsson, 2012), de facteurs géographiques tels que la qualité et l'altitude des sols (Michalopoulos, 2012) et, bien sûr, des conquêtes de territoires, de la

colonisation et de la formation de nouveaux États intégrant différents groupes ethniques. Dans le cas de l'Afrique, Michalopoulos et Papaioannou (2016) soulignent que la colonisation européenne et le tracé arbitraire des frontières sont l'une des principales causes de la polarisation ethnique et des conflits sur le continent.

Un autre volet de la littérature s'est penché sur les facteurs et les politiques qui pourraient favoriser une identité nationale. L'une des dimensions qui a reçu le plus d'attention est le système éducatif (Alesina et Giuliano . Par exemple, Bandiera et autres (2019) fournissent des preuves que les États américains ont introduit des lois sur la scolarité obligatoire pour diffuser les valeurs civiques au sein de la population immigrée. Cantoni et autres (2017) examinent l'impact d'une réforme des programmes scolaires en Chine, et constatent qu'elle a façonné les préférences politiques mais n'a pas eu d'effet significatif sur l'identité ou le comportement national. Clots-Figueras et Masella (2013) montrent que le fait d'être exposé à une réforme introduisant le catalan obligatoire dans les écoles augmente l'identification et les préférences des Catalans par rapport à l'autonomie et à la sécession dans la région. Il existe également quelques cas de politiques similaires qui n'ont pas réussi ou qui ont même eu des conséquences involontaires. Par exemple, Chen, Lin et Yang (2018) ont analysé les effets d'une réforme des programmes scolaires à Taïwan et ont constaté que si cette réforme a renforcé l'identification nationale, les effets ont disparu avec le temps. Fouka (2017) montre que l'interdiction de l'allemand dans les écoles américaines a conduit à un renforcement de l'identité de groupe chez les immigrés allemands.

D'autres recherches ont examiné des politiques de construction de la nation plus radicales, telles que les réinstallations massives de population visant à mélanger des personnes d'origines différentes (Bazzi et al., 2019). Certains travaux ont documenté la puissante force de la propagande, soit en encourageant les préjugés contre des groupes ethniques particuliers (Voigtländer et Voth, 2015), soit en réduisant le sentiment d'identité de son propre groupe (Blouin et Mukand, 2019). Or, les sentiments et les valeurs identitaires peuvent également être influencés par des menaces ou des situations extérieures qui provoquent des chocs émotionnels, en particulier celles impliquant des expériences collectives (Dell et Querubin, 2018 ; Depetris-Chauvin et al., 2020). Par exemple, Depetris-Chauvin et al. (2020) étudient l'impact des succès sportifs nationaux sur l'identité et les conflits dans le contexte de l'Afrique subsaharienne. Ils constatent que les victoires de l'équipe nationale de football dans les compétitions internationales peuvent accroître l'identification nationale (au détriment de l'identité ethnique), conduire à une plus grande confiance interethnique et réduire la prévalence des conflits dans les mois qui suivent. Ils suggèrent que leurs effets sont déterminés par l'expérience collective des grands événements sportifs, bien que les canaux particuliers ne soient pas entièrement compris.

Le deuxième chapitre de cette thèse s'appuie sur les travaux précédents qui mettent en relation les expériences émotionnelles collectives et l'identité nationale, en essayant d'éclairer les mécanismes de cette connexion et ses conséquences politiques potentielles. Pour ce faire, il se concentre sur le mouvement d'indépendance en Catalogne et sur l'influence du Football Club Barcelona (FCB) au cours de la dernière décennie en tant que cadre idéal. Il examine en particulier si les performances du club peuvent avoir un impact sur l'identité catalane, les préférences politiques et, en fin de compte, les résultats électoraux. L'analyse présentée

imite une expérience en ce sens que les individus sont exposés de manière aléatoire à un choc en fonction de la date à laquelle ils sont interrogés. En d'autres termes, je compare les personnes interrogées juste avant ou après un match, en fonction de son résultat. Grâce à cette méthodologie, je peux déterminer l'impact causal du choc émotionnel associé aux défaites/triomphes du FCB sur les attitudes des individus.

Mon travail contribue également à une vaste recherche sur les conséquences des événements sportifs et les performances des équipes nationales et locales. Des recherches antérieures ont montré que, dans une variété de sports, les performances des équipes locales peuvent avoir une incidence sur la criminalité (Munyo et Rossi, 2013), la violence domestique (Card et Dahl, 2011), les résultats des étudiants (Lindo et al., 2012), les décisions judiciaires (Erenand Mocan, 2018) et même la santé des nourrissons (Duncan et al., 2017). Certaines de ces études ont suggéré que les émotions peuvent déclencher des changements de comportement qui pourraient expliquer en partie ces résultats (Van Winden, 2015 ; Passarelli et Tabellini, 2017). Dans le chapitre, j'explore plusieurs mécanismes de ces lignes et je fournis des preuves qu'en effet, les performances du FCB affectent l'euphorie et l'optimisme, ce qui se traduit par une plus grande confiance dans le gouvernement, une identité de groupe plus forte et, par conséquent, des préférences pour plus d'autonomie.

Notez qu'il existe plusieurs différences entre mon cadre et ceux examinés dans des travaux précédents, qui peuvent contribuer à notre compréhension de la façon dont les identités et les préférences sont façonnées dans le monde. Premièrement, la plupart des études réalisées jusqu'à présent ont porté sur des pays en développement présentant un niveau élevé de diversité et de conflits interethniques. Dans mon cas, j'étudie une démocratie consolidée avec des centaines d'années d'histoire et, bien qu'il y ait eu une augmentation des troubles sociaux, il n'y a pas eu d'épisodes récents de conflit. Par conséquent, mes conclusions peuvent être plus facilement extrapolées à d'autres démocraties présentant de multiples réalités "nationales". Deuxièmement, j'ai examiné un club qui a été de plus en plus explicite sur les opinions politiques qu'il adopte, servant d'exemple des conséquences de la politisation du sport.

Segregation et bien-être

Le pays dans lequel on est né est le facteur le plus important qui détermine les perspectives économiques d'un individu. Les écarts de revenu par habitant entre les pays les plus riches et les plus pauvres du monde dépassent un ratio de cent. On pourrait penser qu'à l'intérieur d'un même pays, et après avoir pris en compte les antécédents familiaux, les individus auraient bénéficié de niveaux de vie similaires, cette vision ne pourrait cependant pas être plus éloignée de la réalité. Le quartier dans lequel un enfant grandit est un déterminant majeur de son bien-être futur. Dans le cas des États-Unis, Chetty et al. (2018) montrent que même en tenant compte des caractéristiques de vos parents, il existe des différences marquées dans les perspectives économiques et de vie selon le quartier où vous avez grandi. Alors qu'un garçon noir issu d'une famille à faibles revenus de Pocatello (Idaho) a 20% de chances d'être incarcéré au cours de sa vie, pour un garçon noir du comté voisin de Logan, les chances ne sont que de 0,7%. Ces différences s'expliquent par une combinaison

de facteurs parmi lesquels les camarades de classe et donc la ségrégation scolaire jouent un rôle prépondérant.

Le troisième chapitre de cette thèse vise à explorer les effets de la ségrégation scolaire sur le bien-être à long terme. À cet égard, il exploite une importante réforme de l'éducation sous la dictature de Pinochet, qui a introduit les bons d'études en 1981. À la suite de cette réforme, en seulement 5 ans, le nombre d'écoles privées a doublé et la part des étudiants inscrits dans ces établissements a rapidement augmenté, passant de 15% à plus de 35%. Comme la plupart des parents aisés étaient plus enclins à envoyer leurs enfants dans des écoles privées, cette augmentation des inscriptions dans le privé s'est accompagnée d'une augmentation significative de la ségrégation scolaire. Ce travail de recherche présente plusieurs contributions, notamment en se concentrant sur l'impact de la ségrégation scolaire sur une large mesure telle que le bien-être subjectif, puisque la réforme a été menée à l'échelle nationale, qu'elle s'intéresse à l'ensemble de la population et pas seulement aux minorités, et enfin qu'elle examine ses effets à long terme, des décennies après la mise en place de la réforme. En outre, il utilise des données administratives sur la scolarisation dans les années 1980 pour mesurer de manière appropriée la réponse au changement de politique, et un plan de recherche comparant les individus nés au même endroit à différentes années pour en déduire les implications causales.

La plupart des publications sur la ségrégation scolaire se sont concentrées sur leur impact sur le niveau d'instruction (Guryan, 2004 ; Johnston et al., 2007 ; Reber, 2010 ; Billings et al., 2014) et sur la criminalité (Kling et al., 2005 ; Weiner et al., 2009 ; Billings et al., 2014). Par exemple, Billings et al. (2014) exploitent la manière dont l'abolition d'une politique de bus et de règles strictes qui augmentent la diversité raciale dans les écoles a affecté les résultats des enfants. Ils montrent qu'à la suite de la réforme scolaire, la ségrégation s'est considérablement accrue parce que les quartiers étaient fortement ségrégués, et que cette augmentation de la stratification scolaire a eu des effets négatifs sur les résultats des garçons aux tests et sur le taux de réussite aux examens et a fait augmenter la criminalité chez les garçons issus des minorités. Dans un travail exploratoire, Echenique et ses collaborateurs (2006) examinent la relation entre la ségrégation raciale à l'école et plusieurs résultats obtenus pendant la jeunesse. Dans ce contexte, la ségrégation semble réduire la probabilité des couples interethniques mais n'a pas d'effet significatif sur le bien-être subjectif. Les limites de cette littérature sont qu'elle ne prend en compte que le cas des États-Unis, que la ségrégation est mesurée uniquement sur la base de l'origine ethnique et qu'elle ne prend en compte que les résultats contemporains ou jeunes. Des recherches récentes ont évalué l'impact à long terme du déménagement dans des quartiers plus riches sur des résultats tels que la santé, le bien-être subjectif et les revenus (Sanbonmatsu et al., 2012 ; Ludwig et al. 2012 ; Chetty et al., 2015). Bien que reposant sur une intervention aléatoire, cette recherche a produit des résultats peu concluants et ne tient pas compte de l'équilibre général.

De nombreuses études ont également examiné l'impact des réformes éducatives et, en particulier, celles liées au choix de l'école comme cause principale de la ségrégation scolaire. Certains documents ont révélé que la flexibilité du choix de l'école tend à accroître la ségrégation et à exacerber les inégalités entre les enfants d'origines sociales différentes. Certains

modèles théoriques, tels que Epple et Romano (1998) et MacLeod et Urquiola (2015), suggèrent que les bons d'études pourraient conduire à une stratification plus importante (tant en termes de milieu familial que de capacité des élèves). Des travaux ultérieurs ont en effet montré que les bons d'études ont tendance à provoquer une ségrégation scolaire plus importante dans divers contextes. Des preuves existent pour les États-Unis (Brunner, Imazeki et Ross, 2010 ; Figlio, Hart et Metzger, 2010 ; Chakrabarti, 2013), la Suède (Böhlmark, Holmlund et Lindahl, 2016), le Chili (Hsieh et Urquiola, 2006) et le Kenya (Lucas et Mbiti, 2012), entre autres.

Si les recherches passées ont analysé l'effet du programme chilien de bons d'études sur les performances et la stratification de l'éducation, on sait peu de choses sur les autres résultats et les conséquences à long terme de la réforme. La littérature tend à trouver des résultats mitigés en ce qui concerne la qualité de l'éducation et les performances des étudiants, mais si tant est qu'il y en ait eu, l'impact a été modeste. L'absence de données appropriées avant et après la réforme et d'un plan de recherche convaincant permettant d'obtenir des estimations causales a empêché de tirer des conclusions solides. En revanche, il semble y avoir un consensus sur une forte augmentation de la stratification scolaire due au programme de bons d'études (Hiseh et Urquiola, 2006 ; McEwan, Urquiola et Vegas, 2008 ; Elacqua, 2012 ; Valenzuela, Bellei et De los Rios, 2013 ; Santos et Elacqua, 2016).

Les recherches présentées dans ce troisième chapitre apportent plusieurs contributions. Premièrement, elle s'intéresse à la ségrégation scolaire plutôt qu'à la ségrégation résidentielle, et sur la base du statut socio-économique plutôt que de la race. Ces distinctions sont importantes car dans de nombreux contextes, les changements de la ségrégation scolaire vont de pair avec les changements de la ségrégation résidentielle et il est difficile de démêler la contribution de chacun séparément. De même, si les minorités ont souvent des revenus inférieurs, cela n'est pas nécessairement vrai. Dans mon analyse, je compare des personnes qui sont nées dans la même municipalité et qui ont donc été exposées au même niveau de ségrégation résidentielle, mais selon l'année de leur entrée à l'école, elles seront exposées à une ségrégation scolaire plus importante en raison de la réforme des bons de 1981. Bien qu'il existe des minorités ethniques au Chili, elles représentent une part beaucoup plus faible que dans la population américaine.

Ensuite, le chapitre se concentre sur un pays en développement où les inégalités et la ségrégation en matière d'éducation sont extrêmes. Le Chili est l'un des pays au monde où le secteur privé est le plus présent dans l'éducation. Plus de 50% des étudiants sont inscrits dans des écoles privées. Étant donné qu'il existe une forte corrélation entre le revenu familial et le type d'établissements fréquentés par les enfants, il n'est pas surprenant que le pays se classe parmi les plus élevés en termes de ségrégation scolaire (OCDE, 2014). L'analyse d'un pays en développement est pertinente car ce sont précisément ces pays qui manquent de ressources éducatives et qui pourraient être plus intéressés à encourager les écoles privées par un système de bons d'études comme l'a fait le Chili dans les années 1980.

Troisièmement, il évalue l'impact potentiel sur un résultat peu étudié et général : le bien-être subjectif. Ces dernières décennies ont vu un nombre croissant de recherches s'appuyant sur des données relatives au bien-être subjectif et leur importance pour les

politiques publiques commence à être reconnue (Dolan, Layard, Metcalfe, 2017 ; Frijters et al., 2020). Les données sur le bien-être subjectif sont particulièrement utiles pour analyser les conséquences des politiques publiques car elles englobent plusieurs dimensions du bien-être en un seul indicateur. De plus, elles peuvent mettre en évidence des changements de bien-être qui ne sont pas perçus en utilisant d'autres données économiques. Dans le contexte de l'étude de la ségrégation, elle devient encore plus pertinente étant donné ses effets potentiels sur des aspects subtils tels que le réseau d'amis et de contacts, la précarité et la satisfaction professionnelles, la santé mentale, etc. Un autre aspect crucial est que pour comprendre pleinement les conséquences des réformes de l'éducation, il faut analyser de manière exhaustive leurs conséquences potentielles à long terme. À cet égard, j'évalue l'impact de la réforme de 1981 trois décennies après qu'elle ait eu lieu, ce qui implique que si le bien-être des individus est toujours affecté, les gains potentiels (coût) des réformes éducatives ne peuvent pas être évalués uniquement à court terme.

Enfin, la plupart des recherches sur la ségrégation scolaire se concentrent sur des contextes particuliers tels qu'une ville ou un comté qui connaît une réforme. Étant donné que ces politiques peuvent avoir toutes sortes de conséquences sur les zones voisines, pour avoir une vision globale de ses effets, il convient de les prendre en compte. Dans mon cas, étant donné la nature nationale de la réforme chilienne, tous les individus ont été touchés et en analysant les données de l'ensemble du pays, je peux prendre en compte toutes les retombées potentielles et les effets d'équilibre général générés par l'augmentation de la ségrégation. Ces derniers sont en effet cruciaux si l'on veut comprendre les conséquences potentielles de l'intensification de politiques similaires.

Plan de Thèse

Cette thèse exploite des expériences quasi-naturelles, historiques et contemporaines, pour fournir des preuves causales sur de vastes questions à la frontière de la littérature sur le développement et l'économie politique. En particulier, elle examine les effets à long terme de l'émigration dans les communautés d'origine, la formation et les moteurs des identités nationales et des préférences politiques, et les impacts à long terme de la ségrégation scolaire sur le bien-être.

L'une des questions les plus importantes concernant la migration est de savoir comment elle affecte les communautés d'origine. En particulier, les économistes ont accordé une attention particulière aux conséquences sur le stock de capital humain (Beine, Docquier et Rapoport, 2008). Aujourd'hui, il est difficile d'établir une relation de cause à effet entre la migration et l'éducation dans les communautés d'origine en raison de l'existence de facteurs affectant à la fois les décisions de migration et les investissements dans l'éducation et de la possibilité d'une causalité inverse (Gibbons, McKenzie et Stillman, 2011). Si les chercheurs ont mis en œuvre diverses stratégies pour répondre à ces préoccupations, les approches proposées intègrent souvent d'autres mises en garde et problèmes de validité externe. En outre, une caractéristique commune de la littérature passée est que les limites des données ont fait que la plupart des études ont adopté une vision à court terme et

analysé uniquement les effets de la migration à très court terme.

Une perspective à long terme est particulièrement pertinente dans le contexte de l'accumulation de capital humain, car l'impact de l'émigration peut facilement changer au fil du temps, tant en termes de taille que de signe. En outre, certains mécanismes ne jouent un rôle qu'à long terme. Comme le soulignent les contributions théoriques et empiriques (McKenzie et Rapoport, 2010 ; Dustman et Weiss,2007), le type de sélection à l'émigration, les incitations à acquérir une éducation et le degré de migration de retour sont tous des facteurs clés pour comprendre les effets de la migration à l'origine, et ils ont une influence à différents stades. Au-delà de cela, les migrants peuvent également avoir un impact à long terme sur leurs communautés d'origine par le biais de mécanismes plus subtils. Par exemple, ils peuvent contribuer au financement de biens publics tels que les écoles ou agir comme des moteurs de changement culturel et contribuer à façonner des normes qui encouragent les investissements dans le capital humain des générations futures.

Le premier chapitre de cette thèse revient sur la question de l'impact de l'émigration sur la formation scolaire à l'origine, avec une approche historique et à long terme. En particulier, il examine l'impact de l'émigration de masse sur l'accumulation de capital humain dans les communautés d'origine tout au long d'un siècle. Pour ce faire, je me concentre sur l'une des périodes où les flux migratoires ont été les plus importants de l'histoire moderne : l'ère des migrations de masse (Hatton & Williamson,1998). Entre 1850 et 1930, plus de 40 millions d'Européens ont traversé l'Atlantique en direction des Amériques. Afin de zoomer et de pouvoir effectuer une analyse causale, je prends comme laboratoire la diaspora galicienne. Au cours de la période 1900-1930, plus de 1,1 million de personnes ont quitté cette région espagnole vers l'Amérique latine, un chiffre représentant environ 60% de sa population de 1900. En termes relatifs, la Galice avait un taux d'émigration plus élevé que celui de tout autre pays européen à l'époque. L'ampleur de cet événement historique, ainsi que les similitudes entre la Galice de 1900-1930 et d'autres pays en développement aujourd'hui, en font un cadre exceptionnel pour étudier les effets à court et à long terme de l'émigration massive.

Je combine des données récemment numérisées provenant de multiples sources historiques (par exemple, des recensements de population, des listes d'embarquement, des associations de migrants, parmi beaucoup d'autres) avec des données administratives et d'enquête modernes pour construire une base de données unique des municipalités galiciennes de 1860 à aujourd'hui. L'une des principales innovations est que je propose et valide un nouvel indicateur de la migration au niveau local : les résidents manquants. Je me fonde notamment sur la proportion d'hommes qui étaient absents de leur foyer le jour du recensement. J'effectue plusieurs vérifications pour confirmer qu'il s'agit d'un bon indicateur du stock d'émigrants.

Pour effectuer une analyse causale, je mets en œuvre un plan de recherche dans lequel j'instrumenterai la part des hommes disparus par deux sources de variation plausibles et exogènes. Pour déduire l'intensité de la migration dans une municipalité donnée, j'utilise un indicateur de l'émigration des pionniers mesuré plusieurs décennies avant l'ère de l'émigration massive. Pour déduire le calendrier des flux migratoires, je m'appuie sur un facteur d'attraction variable dans le temps au niveau local. Plus précisément, comme la

plupart des pionniers de l'émigration étaient des hommes, j'utilise le sex-ratio adulte mesuré en 1860 comme indicateur de ces premiers flux. Comme facteur d'attraction, je construis une mesure décennale de la croissance du PIB par habitant à la destination moyenne pour chaque municipalité en utilisant des données sur la part des individus qui émigrent vers chaque pays. En utilisant cette stratégie empirique, j'analyse les moteurs de l'émigration au niveau local, ses effets à court et moyen terme sur les taux d'alphabétisation (1900-1930), et son impact à long terme sur différentes mesures du niveau d'éducation (1930-1901). En outre, j'utilise d'autres sources de données pour explorer plusieurs mécanismes, notamment le rôle joué par les associations de migrants à l'origine et l'effet potentiel des migrants en tant que moteur de changement des valeurs.

Je constate que l'émigration a eu un effet négatif important à court terme sur le taux d'alphabétisation à l'origine. Cet effet est présent tant pour les hommes que pour les femmes et s'explique par le fait que les migrants ont été sélectionnés positivement, c'est-à-dire (avec un taux d'alphabétisation plus élevé que la moyenne de la population). Aujourd'hui, une décennie plus tard, l'impact de l'émigration devient positif et significatif, ce qui signifie qu'elle a entraîné des gains nets dans les taux d'alphabétisation. Ces effets positifs ne sont toutefois présents que pour les hommes. Je discute de plusieurs canaux qui pourraient expliquer ces résultats. Premièrement, plus de 60% des migrants retournent en Galice, ce qui compenserait en partie l'effet de composition à court terme. Deuxièmement, l'émigration semble avoir augmenté le taux d'alphabétisation des jeunes enfants.

J'analyse ensuite les conséquences à long terme de la diaspora galicienne sur le capital humain à l'origine. En utilisant les données par cohortes du recensement contemporain, je constate que l'émigration a conduit à des taux plus élevés d'achèvement du cycle primaire chez les personnes nées dans les années 1930, 1940 et plus tard. Je constate également des effets positifs sur l'achèvement du cycle secondaire pour certaines cohortes et sur l'achèvement des études universitaires dans le dernier recensement. Ces résultats soulignent l'importance de considérer de longs horizons et plusieurs résultats pour avoir une image claire des effets de l'émigration à l'origine.

J'explore également de nouveaux mécanismes pour fournir des preuves de la persistance à long terme du capital humain. Je documente d'abord que les Galiciens ont financé des centaines d'associations à but éducatif qui ont financé la construction d'écoles dans leurs villes natales. En combinant les données sur ces écoles et les données de recensement par cohortes, je montre que ces investissements ont permis d'améliorer les résultats scolaires des enfants qui y sont exposés. Des preuves anecdotiques suggèrent que les migrants et leurs associations ont été des facteurs actifs de diffusion de l'information et des normes sur l'importance de l'éducation. J'utilise des données d'enquête dans lesquelles les individus indiquent leur évaluation de l'éducation et de l'effort pour faire la lumière sur ces canaux. Je constate que les individus des municipalités les plus exposées à l'émigration un siècle plus tôt aujourd'hui affichent une plus grande valorisation de l'éducation.

Les conclusions de ce chapitre ont des implications importantes pour la recherche sur les migrations et les politiques publiques. Tout d'abord, elles soulignent la pertinence d'adopter une vision à long terme, car l'impact de la migration et les personnes concernées

peuvent évoluer dans le temps de manière inattendue. Par exemple, alors qu'à court terme, l'émigration a fait baisser le niveau d'alphabétisation des hommes et des femmes en raison de la sélection, les décennies suivantes n'ont eu un impact positif que sur le niveau d'éducation des hommes, et à long terme, les hommes et les femmes ont bénéficié de la même manière. Ils soulignent également la nécessité de proposer et d'évaluer d'autres mécanismes que les transferts monétaires, tels que les transferts collectifs sous forme d'investissements dans les biens publics et les transferts sociaux. Mes résultats suggèrent qu'encourager l'engagement des migrants dans leurs communautés d'origine (par exemple, par le biais d'associations locales, de campagnes d'information, de collectes de fonds) peut être une stratégie prometteuse pour promouvoir l'accumulation de capital humain et le développement économique à l'origine. Des politiques publiques innovantes pourraient être mises en œuvre pour renforcer ces liens et encourager la transmission de transferts de fonds collectifs visant des objectifs spécifiques. Enfin, mes conclusions jettent un nouvel éclairage sur l'impact de la migration sur l'adoption de nouvelles normes culturelles, un canal qui peut entraîner des effets persistants à long terme. Les travaux futurs devraient examiner plus en profondeur l'interaction entre ces différents mécanismes et explorer l'impact à long terme de l'émigration sur d'autres aspects sociaux et économiques dans les communautés d'origine.

Le deuxième chapitre explore le lien entre l'identité et les préférences politiques, et la manière dont elles peuvent répondre à des chocs émotionnels. Il existe une abondante littérature soulignant comment les clivages identitaires peuvent être une source majeure d'instabilité politique et économique (Besley et Reynal-Querol, 2014). En réponse à cela, les États se sont historiquement engagés dans des politiques de construction de la nation, favorisant un sens commun de l'identité par le biais d'une scolarisation de masse et de réformes éducatives (Bandiera et al., 2019 ; Cantoni et al., 2017 ; Clots-Figueras et Masella, 2013), de la propagande et de l'utilisation des médias (Blouin et Mukand, 2019), et même des réaffectations de masse (Bazzi et al., 2019). De même, les chocs émotionnels, en particulier ceux qui impliquent des expériences collectives, peuvent également façonner l'identité et les valeurs de manière complexe (Dell et Querubin, 2018 ; Depetris-Chauvin et al., 2020).

À cet égard, l'une des forces les plus puissantes qui animent les passions et les sentiments des individus est le sport. Le football, en particulier, a souvent été associé à la construction d'identités ethniques et nationales, avec d'innombrables exemples de clubs encourageant et embrassant certaines identités (Foer, 2004 ; Shobe, 2008a,b ; Úbeda-Colomer et al., 2017) Dans un contexte de clivages forts, ce cadre de rivalité, d'opposition de positions, alimenté par la composante émotionnelle du sport, peut intensifier les différences entre deux entités dans un contexte de clivages forts (Ramonet, 1998), comme l'illustrent les conflits qui avaient été alimentés par les rivalités du football (Úbeda-Colomer et al., 2017).

Ce document examine si un lien étroit entre un club sportif et un sentiment d'identité peut influencer l'identité nationale, les préférences politiques et conduire à un changement politique. Je me concentre sur le cas de la Catalogne et du Football Club Barcelona (FCB), l'un des clubs les plus réputés au monde, qui a historiquement adopté et promu l'identité catalane. En fait, l'hypothèse selon laquelle le succès du FCB pourrait revigorer les sentiments pro-sécession et renforcer le processus d'indépendance catalan avait déjà été

proposée par certains intellectuels dès les années 1990 (Vázquez-Motalbán, 1999).

Les principales données utilisées proviennent de deux enquêtes d'opinion menées régulièrement en Espagne par le CEO (Centro d'Estudis d'Opinó) et le CIS (Centro de Investigaciones Sociológicas). Le baromètre du CEO interroge environ 1500 personnes en Catalogne chaque trimestre tandis que le baromètre du CIS interroge environ 2500 personnes en Espagne chaque mois. Les deux enquêtes comportent une question d'identité dans laquelle les personnes doivent se décrire comme "uniquement espagnol", "plus espagnol que catalan", "aussi espagnol que catalan", "plus catalan que espagnol" ou "uniquement catalan". Ils interrogent également les individus sur leurs préférences concernant la décentralisation de l'État et la possibilité pour les régions de devenir indépendantes. Je fais correspondre les données de l'enquête avec des informations détaillées sur les matchs et les trophées du FCB en fonction de la date des entretiens. Étant donné les différences entre les enquêtes du CEO et du CIS, je les utilise pour mener deux analyses distinctes mais complémentaires : l'une portant sur l'impact à court terme des matchs du FCB et l'autre sur l'effet à court terme des titres du FCB à la fin de la saison.

Afin d'arriver aux effets causaux des matchs du FCB ou des titres du FCB, j'exploite les chronologies exogènes des interviews autour des événements. En ce sens, le fait qu'un individu soit interrogé juste avant ou après la défaite est aléatoire. Il en va de même pour les personnes interrogées au cours d'une année où le FCB remporte plus ou moins de titres. J'effectue plusieurs analyses qui soutiennent l'hypothèse selon laquelle les individus *traités* et *control* sont comparables sur les observables.

En exploitant les données de l'enquête avec des informations sur les matchs et les trophées du FCB entre 2009 et 2019, je constate que les performances du FCB entraînent de brusques changements (temporaires) dans l'identification des Catalans, les préférences en matière de décentralisation et le soutien à la sécession en Catalogne. Personnes interrogées un jour après une défaite sont 11% moins susceptibles de faire état d'un fort sentiment de catalan et sont 15% moins susceptibles de soutenir l'indépendance de la Catalogne. En revanche, les Les titres de FCB en fin de saison permettent d'identifier et de soutenir les l'autodétermination d'environ 11%. Ces effets persistent pendant quelques semaines avant de disparaître. Mes résultats sont conformes aux conclusions de Depetris-Chauvin et al. (2020) dans le contexte du football africain, mais les implications sont cependant contraires : le football peut contribuer à la fois à unir et à désunir un pays. En exploitant les données sur les supporters du FCB dans les municipalités, je trouve des preuves que les défaites de matchs juste avant les élections ont un impact significatif sur la diminution de la participation politique.

J'effectue plusieurs analyses d'hétérogénéité et j'explore les mécanismes potentiels de ces effets. En particulier, j'examine les indicateurs des changements d'humeur, comme la satisfaction de vivre, l'optimisme, la confiance dans les hommes politiques, etc. Je constate que les titres FCB entraînent des changements d'optimisme, améliorent les perspectives économiques et politiques des individus, et conduisent également à moins de méfiance à l'égard du gouvernement. Ces résultats confirment l'idée selon laquelle les changements d'identité et de préférences sont dus à des modifications à court terme de l'euphorie.

Les conclusions de ce chapitre ont des implications politiques importantes. Tout d'abord, elles montrent que les croyances et les préférences des individus peuvent être affectées par des chocs aléatoires par le biais de changements d'humeur. En outre, malgré la nature temporaire des changements d'identité et de préférences provoqués par le football, ils pourraient se traduire par des changements permanents en affectant les résultats électoraux. Dans l'ensemble, les résultats appellent à séparer autant que possible le sport de la politique, et à être prudent lors de la prise de décisions à haut enjeu car elles peuvent être influencées par des chocs émotionnels.

Le troisième chapitre de cette thèse examine les effets à long terme de l'exposition à la ségrégation scolaire pendant l'enfance. Un grand nombre d'études ont examiné les conséquences de la ségrégation scolaire pour de nombreuses dimensions telles que le niveau d'instruction (Guryan, 2004 ; Johnston et al., 2007 ; Reber, 2010 ; Billings et al., 2014) et la criminalité (Kling et al., 2005 ; Weiner et al., 2009 ; Billings et al., 2014). Pourtant, il n'existe pas de preuves empiriques de ses conséquences potentielles à long terme sur le bien-être.

Pour étudier les effets à long terme de la ségrégation scolaire, je profite d'un événement historique unique qui a façonné le système éducatif chilien pendant des décennies. En 1981, le gouvernement militaire de Pinochet a introduit un système de bons d'études qui permettait aux enfants de fréquenter des écoles privées subventionnées sans avoir à payer de frais de scolarité. En conséquence, entre 1981 et 1988, le nombre d'écoles privées a doublé et la proportion d'enfants fréquentant ce type d'établissements est passée de 15 à 35%. Pourtant, cet exode a été caractérisé par des étudiants des classes moyennes et supérieures, ce qui a entraîné une forte augmentation de la ségrégation scolaire.

J'utilise l'administration des inscriptions dans les municipalités avant et après la réforme pour obtenir une mesure précise de l'ampleur de la privatisation au cours des années 1980. Étant donné le lien étroit entre le revenu des parents et la probabilité que les enfants fréquentent une école privée subventionnée, j'interprète l'augmentation des inscriptions dans le privé entre 1981 et 1988 comme un indicateur de la montée de la ségrégation à l'époque. Je relie ensuite les données administratives des inscriptions aux données détaillées des enquêtes individuelles pour l'ensemble du pays pour les années 2011 et 2013, en fonction de la municipalité de naissance des individus. Il est important de noter que les données d'enquête contiennent une question standard de bien-être subjectif qui est au centre de l'analyse. Les individus sont invités à évaluer leur satisfaction dans la vie au cours des dernières semaines sur une échelle de 1 à 10.

Pour identifier l'effet causal de la réforme de 1981, je réalise une analyse qui exploite à la fois la variation entre les cohortes et l'espace. En bref, je compare des individus nés au même endroit mais à des années différentes, ce qui détermine leur exposition à la réforme. Les individus nés au début des années 1970 et donc scolarisés dans les années 1980, alors que la ségrégation était en hausse, sont définis comme le groupe traité. Les individus nés au début des années 1960, et donc déjà sortis de l'enseignement obligatoire en 1981, constituent le groupe témoin ou contrefactuel. J'examine ensuite comment les écarts de

satisfaction dans la vie entre ces deux groupes sont influencés par le degré de privatisation des écoles au cours des années 80. J'utilise également les informations sur l'éducation des parents pour classer les individus en fonction de leur milieu familial et donc de leur probabilité d'avoir fréquenté une école privée.

Je trouve des preuves que la réforme a eu des effets durables sur la satisfaction de la vie et qu'elle a accru les inégalités. Parmi les personnes issues de milieux familiaux plus défavorisés, une augmentation de 10 points de pourcentage des inscriptions privées subventionnées lorsqu'elles étaient jeunes est associée à une baisse de la satisfaction de la vie adulte dans 5,9% d'un déviation standard.. Cet impact équivaut à l'effet d'une diminution du revenu des ménages d'environ 30% ou d'une réduction de l'éducation de 3 ans. En revanche, les personnes issues des milieux les plus favorisés semblent avoir bénéficié de la réforme, bien que les preuves soient moins solides. L'impact diffère sensiblement selon le sexe et les schémas migratoires. L'effet négatif sur les personnes d'origine familiale faible est entièrement dû à celles qui sont restées dans leur municipalité de naissance. L'effet positif sur les personnes d'origine familiale élevée est plus important et significatif chez les hommes et les personnes qui ont migré à l'intérieur de leur propre pays vers d'autres municipalités. Si l'on se concentre sur les personnes qui ont déménagé, les personnes d'origine familiale modeste qui se retrouvent dans des zones fortement privatisées sont nettement moins bien loties. Dans ce cas, une augmentation de 10 points de pourcentage des inscriptions privées est associée à une réduction de la satisfaction de la vie adulte de 11,1% d'un déviation standard.

J'effectue un certain nombre d'analyses complémentaires pour explorer la robustesse des résultats. Par exemple, je modifie les fenêtres de cohorte considérées comme des groupes traités et des groupes témoins comme si la réforme avait eu lieu une année différente et j'évalue également des mesures alternatives d'exposition en tenant compte du nombre d'années pendant lesquelles les individus ont été scolarisés après la réforme. J'explore également les effets potentiels sur d'autres résultats tels que la scolarisation, les revenus et la santé, sans trouver de preuve que la réforme a eu un impact à long terme sur ces dimensions.

Les conclusions de ce chapitre illustrent comment une politique destinée à bénéficier aux élèves les plus pauvres en leur donnant la possibilité de fréquenter de meilleures écoles a eu des effets pernicieux en raison de sa mauvaise conception. Le choix n'a jamais été une véritable option pour les familles les plus vulnérables et, par conséquent, elles n'ont pas pu bénéficier efficacement de la réforme. Le fait que ce changement de politique ait eu un effet permanent sur la satisfaction de la vie adulte exige une conception bien pensée des politiques éducatives, sinon, même les mesures bien intentionnées peuvent compromettre le bien-être des personnes pour le reste de leur vie. Cela remet également en question l'idée selon laquelle les politiques éducatives ne devraient être évaluées qu'en fonction de leur impact sur l'apprentissage des enfants. Les années d'éducation sont essentielles pour la formation des identités, des croyances et des valeurs, des réseaux de pairs, etc., des aspects qui peuvent ne pas être entièrement malléables par la suite. Les recherches futures devraient explorer en profondeur les mécanismes qui sous-tendent l'effet négatif sur les individus issus de familles défavorisées afin de mettre en lumière la manière dont ces effets auraient pu être évités.

Chapter 1

Mass Emigration and Human Capital over a Century: Evidence from the Galician Diaspora

Mass Emigration and Human Capital over a Century: Evidence from the Galician Diaspora [*]

Abstract

This article examines the effects of emigration on human capital accumulation at origin throughout a century. The context is the Galician diaspora, a massive episode with the equivalent of 60% of the region 1900's population emigrating to Latin America during 1900-30. I construct a database of all Galician municipalities combining newly-digitized historical data with contemporary census and survey data and exploit two sources of plausibly-exogenous variation for identification: pioneer emigration caused by extreme rainfall and changes in economic growth in the main migrant destinations. I find that while emigration depressed literacy rates at origin in the short run, its impact became positive after one decade and led to gains in human capital that still persist one hundred years later. I provide evidence of two novel mechanisms on how emigrants raise human capital in the long run. Galician emigrants funded associations that financed the construction of schools in their hometowns and diffused norms conducive to a persistent change in beliefs about the value of education and effort.

Keywords: Mass emigration, Human capital, Migrants' associations, Values, Long-run.
JEL Classification: I25, J11, O15, N33, N34.

1 Introduction

During the Age of Mass Migration (1850-1930), over 40 million Europeans crossed the Atlantic towards the Americas in one of the largest migration flows in modern history (Hatton & Williamson, 1998). This unique episode led to profound transformations in both continents whose effects can still be seen today. Recent research has examined the long-term consequences of migration for receiving or host countries, showing that it promoted economic development and higher levels of human capital (Rocha, Ferraz and Soares, 2017; Droller, 2018; Sequeira, Nunn and Quian, 2020). The implications for the sending communities however remain much less unexplored (Karadja and Prawitz, 2019; Anderson, Karadja and Prawitz 2020). Specifically, there is still no empirical evidence on the long-run effects of emigration on human capital accumulation at origin.

A long-term perspective is especially relevant in the context of human capital accumulation since the impact of emigration and the mechanisms involved may change over time. As underlined by the empirical and theoretical literature (McKenzie and Rapoport, 2010; Dustman and Weiss, 2007), the type of selection into emigration, the incentives to acquire education and the extent of return migration are all key factors that play a role at different stages. Moreover, migrants may also have a long-run impact on their communities of origin through more subtle mechanisms. For instance, they may contribute to the financing of public goods or the shaping of norms and values that encourage investments in human capital of the future generations.

This paper examines the impact of mass emigration on human capital accumulation at origin throughout an entire century. I study one of the greatest emigration episodes of modern history in Europe: the Galician diaspora. Over the 1900-1930 period, more than 1.1 million people left this Spanish region towards Latin America, a figure representing 58% of its 1900 population. In relative terms, Galicia had higher emigration rates than those in any other European country at the time. The magnitude of this historical event, together with the similarities between Galicia in 1900-1930 and other developing countries today, make it an exceptional setting to study the short and long-run effects of mass emigration.

In this work, I combine newly digitized data from multiple historical sources (e.g., population censuses, embarkation lists and migrants' associations) with modern administrative and survey data to construct a unique database of Galician municipalities from 1860 to date. I use data on absent men as a novel proxy for emigration and instrument this via a measure of migrant networks (pioneer emigration caused by extreme rainfall) and a pull factor (economic growth at destination).

Using this empirical strategy, I analyze the drivers of emigration at the local level, its short and medium-run effects on literacy rates (1900-1930), and its long-run impact on different measures of educational attainment (1930s-2010s). In a nutshell, I find that while emigration depressed literacy rates in the short run, one decade later the effect became positive and led to gains in human capital that still persist around one hundred years later. Using data on migrants' associations and their investments at origin as well as census and survey data, I provide evidence that this long-term persistence is partially due to migrants' financing of new schools in their hometowns and a change in beliefs about the value of education and effort.

A key novelty of my study is to propose and validate a new proxy for historical emigration based on the concept of missing residents.[1] The population censuses of various countries (e.g., Spain, Italy, and Portugal) gathered information of all household members, including those temporarily absent.[2] I perform an exhaustive number of checks that validate the share of missing men as a good proxy for emigration. For instance, I show that the numbers of missing men and sea departures in Galicia are extremely similar, and follow the exact same trend between 1877 and 1930. I also find no evidence that these missing men could reside somewhere else in Galicia or Spain.[3]

Analyzing the effects of historical emigration on human capital accumulation is a challenge for a number of reasons. First, historical data can suffer from substantial measurement error, leading to attenuation bias. Second, there may be confounding factors affecting both migration flows and investments in education. Last, changes in education may influence migration decisions leading to reverse causality. To deal with these issues, I implement an empirical strategy in which I instrument the share of missing men by two sources of plausibly exogenous variation. In the historical analysis using decadal panel data between 1900 and 1930, I instrument the share of missing men in a given year by a time-varying pull factor and its interaction with a proxy for pioneer migrant networks (measured four decades before the onset of mass emigration). In addition, I control for municipality time-invariant and aggregate time-varying unobserved factors that capture most of the variation in missing men and literacy. When examining the long-run effects of emigration using a cross-section of municipalities, I instrument the average share of missing men during

---

[1]The study of European mass emigration has traditionally been limited by the lack of local level data. In Spain, for instance, official migration records only exist at the province level (49) and for a handful of years. Information on missing residents however is available for all municipalities and census years since 1877.

[2]Household members who did not sleep in the dwelling in the past two nights prior to the census were registered as missing regardless of the duration or reason of their absence. Since most Galician migrants were young men departing alone, household heads would report them as missing. Note that the census took place on December 31st precisely to minimize absences others than those due to emigration.

[3]One additional advantage of relying on data of missing men is that it captures clandestine emigration (i.e., overlooked by official statistics) that at the time represented 20% of recorded departures in Spain (Sánchez-Alonso, 1995).

1900-1930 by the proxy for migrant networks (1860) and control for other historical factors.

The main ingredient of my identification strategy is a proxy for pioneer migrant networks.[4] I rely on the fact that the first emigration waves (1840-1860) were largely driven by extreme rainfall that accentuated agricultural crises (Rodríguez Galdo, 1977; Hernández Borge, 1986; Eiras Roel, 1991). These pioneers fostered new departures in the following decades and were a key force during the mass emigration era (1900-1930). Since most pioneers were men, I rely on the adult sex ratio in 1860 across municipalities as proxy for the size of migrant networks.[5] As a pull factor, I use the variation in economic growth across destinations during each decade. In particular, I construct a measure of GDP per capita growth in the average destination country for every municipality using historical records on the destination of over 17,000 migrants during 1900-1930.

To assess the validity of my identification strategy, I perform several robustness checks. First, I show that the adult sex ratio in 1860 is uncorrelated with the evolution of relevant factors between 1860 and 1900, such as population or literacy rates. Second, I show that economic growth at destination and origin are completely uncorrelated and pay special attention to the concern that other local economic shocks (e.g., international trade and weather shocks) may confound my results. Third, I use alternative weights to construct my pull factor and exploit different measures of migrant networks.[6] Finally, I control flexibly for a large set of historical controls and perform multiple sample restrictions. Overall, the results remain robust to all these checks.

I begin my analysis by studying the drivers of emigration at the local level. Using both a decadal panel of municipalities between 1900 and 1930 and a cross-section, I find that the proxy for migrant networks is the major determinant of emigration. The adult sex ratio in 1860 explains 20% of the variation in the average share of missing men during 1900-1930. Higher economic growth at destination is associated with increases in the share of missing men, especially in areas with a larger migrant network. This mediating effect of migrant networks is consistent with other findings in the literature (McKenzie and Rapoport, 2010; Brum, 2019). I also find suggestive evidence that migrants were credit constrained, as rainfall shocks correlate negatively with changes in missing men. Overall, the instruments are highly significant and robust across all the specifications.

I then analyze the short and medium-run effects of emigration on literacy rates during the mass emigration era (1900-1930). Exploiting variation in the pull and network instruments, I find that

---

[4]I refer to migrant networks as a group of migrants from the same origin community and residing in the same place.

[5]I cannot rely on missing men in 1860 because this information is only available from 1877 onwards.

[6]For instance, I use the distance to the closest transatlantic port in a similar fashion to Karadja and Prawitz (2019), but taking into account the road network of Galicia in the nineteenth century. This variable reflects the diffusion of information and transportation costs. Additionally, I use the raw data of rainfall variability during 1840-1860.

emigration depressed literacy rates in the short run. On average, a one standard deviation rise in the share of missing men led to a 51% of a standard deviation drop in literacy rates. This negative impact however switches sign one decade later. Introducing the share of missing men one decade earlier as the main explanatory variable, I find that the medium-run effect of migration is positive and significant for men and zero for women. I discuss potential mechanisms that may explain these effects. First, I present data on migrants' characteristics and suggest that the initial drop in literacy can be partly explained by the departure of people who were more educated than the average. Second, the medium-run effects could be accounted for greater investments in children's education and return migration. Using literacy information by cohort, I find suggestive evidence that i) emigration increased literacy rates for both boys and girls, ii) there was no negative selection into return migration and iii) men were much more likely to return than women.

I next examine the long-run effects of mass emigration on human capital accumulation. Instrumenting the average share of missing men (1900-1930) with the migrant network measured in 1860, I find that historical emigration led to persistent gains in human capital (1981-2011). A one standard deviation rise in the average share of missing men (i.e., 7 percentage points) is associated with a 50% of a standard deviation rise in the share of individuals with completed secondary education in 2011 (i.e., 5 percentage points). In line with the medium-run effect in the historical analysis, the impact for women is delayed and becomes positive and significant only in the 2000s. I find similar positive effects, but smaller in magnitude, when looking at the share of college-educated individuals. To see whether these effects can be driven by internal migration, I separate the sample between individuals who were born in the municipality and those who moved there. There turn out to be no significant difference in the results for these two groups. I also examine other contemporary outcomes and find no impact on the number of schools or libraries, but a positive impact on the number of local associations with an educational goal.

After establishing that mass emigration led to gains in human capital in the long run, I explore two potential mechanisms for this long-term persistence: the role of migrants' investments at origin and social remittances. Galician migrants founded local associations with the primary goal of financing the construction of schools in their hometowns, and founded over 300 new schools between 1910 and 1940 (14% of the stock of public schools in 1908). I collected and assembled data on hundreds of these migrants' associations with detailed information on the country and year of foundation, the Galician municipality to which they are related, and the exact location and year of opening of the schools they financed. I then merge this database with education data by birthplace

and cohort from the 1981 census. A cohort analysis reveals that migrants' schools had a significant and positive impact, increasing the rates of primary education completion for the children young enough to benefit from them. Consistent with historical narratives, I find that migrants' associations seem to have encouraged further migration. However, it is important to note that since most associations financed several schools, the net effect on educational attainment is positive.

To better understand the motivations of migrants to found associations and invest in their home-towns, I leverage historical data from different sources and perform a cross-sectional analysis. I first find that the size of the migrant community seems to have played a crucial role. A one standard deviation increase in the average share of missing men (1900-1930) is associated with one additional migrant association and 1.6 more migrant schools in the municipality of origin. Migrants were more likely to found an association and finance a new school if there was already a greater need for them to begin with. I also find that rainfall shocks (i.e., negative income shocks) are negatively and strongly correlated with the number of associations and schools founded, suggesting that migrants faced a trade-off between sending individual and collective remittances.

Another channel by which migrants could have fostered higher investments in human capital is social remittances related to education. Anecdotal evidence suggests that migrants realized the importance of education once in Latin America and transmitted these values in their letters, through the role of migrants' association and directly after their return. To test this hypothesis, I merge the historical data with contemporary survey data on beliefs and values. Using the proxy for migrant networks (1860) as instrument, I find that historical emigration (1900-1930) led to greater value being put on education and effort today. This result is robust to controlling for other historical and contemporary factors. Additionally, I find suggestive evidence that these effects are stronger in municipalities that were connected to Latin American countries where education was more valuable. I find no changes in other individual beliefs such as the degree of satisfaction with public education or the willingness to pay higher taxes to invest more in public goods.

My paper relates to a number of different literatures. First, it contributes to the growing research exploring the long-run effects of migration. Recent work has documented positive long-run effects on economic development and human capital accumulation in the areas receiving migrants (Rocha, Ferraz and Soares, 2017; Droller, 2018; Sequeira, Nunn and Quian, 2020). Regarding the impacts at origin, the literature has found positive effects on innovation (Anderson, Karadja and Prawitz, 2020), and mixed results in terms of economic development and human capital formation (Dinkelman and Mariotti ,2016; Testa, 2020). I provide novel evidence that, despite a short-run

loss, the Galician mass emigration led to gains in human capital that persisted across several generations. To understand the variety of results, it is crucial to consider the context of each study. Dinkelman and Mariotti (2016), for instance, examine the circular migration from Tanzania to South African mines, which for political reasons ended abruptly forcing all migrants to return home. This is a setting particularly beneficial for a positive impact on human capital formation given that all migrants had to return and they kept all their earnings. In the case of Testa (2020), he analyzes an episode of forced migration (Germans expelled from Czechoslovakia after WWII) implying that parts of the population were expelled from their hometowns and not allowed to come back. In this regard, the Galician diaspora provides a more general setting to understand the effects of emigration since the decisions to emigrate, return, or send remittances, were not influenced by institutional factors or wars. Moreover, it was similar to other migration flows from Southern Europe during the age of Mass Migration and to other migration flows today, characterized by male migrants often leaving their family behind, and thus, maintaining strong links with their communities of origin in clear contrast with episodes of forced migration (Becker and Ferrara, 2019).

Second, I contribute to the literature on the effects of migration on the education of the left-behind and the debate on the brain drain from a historical perspective.[7] The research on the Age of Mass Migration has pointed to both negative (Abramitzky, Boustan and Eriksson, 2012; Connor, 2019) and positive (Spitzer and Zinram, 2018) selection of migrants depending on the context. My results suggest that Galician migrants were positively selected and although their departure reduced literacy rates in the short run, the high levels of return migration (in line with other countries, Bandiera, Rasul and Viarengo, 2013) and greater investments on children's education led to net gains in human capital. This is the first evidence of positively selected migration followed by brain gain in the context of the Age of Mass Migration.

My paper also relates to a recent literature on migration and culture. For instance, Knudsen (2019) shows that migrants from Scandinavian countries were on average more individualistic and, as a result, their departure raised the level of collectivism at origin.[8] Becker et al. (2020) show that Polish forced migrants subsequently became less materialistic and invested more in education. I contribute to this literature by examining how emigration affected beliefs about the value of ed-

---

[7]See for instance McKenzie and Rapoport (2006, 2010), Antman (2012), Batista, Lacuesta and Vicente (2011), Cortes, Gibson and McKenzie (2011, 2012), and Shresta (2016). Docquier and Rapoport (2012) review the existing literature.

[8]In a related work, Bazzi, Fiszbein and Gebresilasse (2020) show that individuals living in areas that were the US frontier for longer periods of time display higher levels of individualism.

ucation and effort. Rather than the result of migrant selection or forced migration, my findings seem to suggest that an exposure to an environment where education is highly valued may make migrants change their norms and transmit these to their communities of origin.

Finally, some work has examined the role of non-governmental institutions that provide public goods and contribute to local development, investments in human capital and cultural change. The evidence ranges from religious institutions centuries ago (Valencia-Caicedo, 2019; Calvi et al., 2019) to contemporary migrant associations (Chavet et al., 2015). I document that Galician migrants established hundreds of associations with the aim of investing in their home communities, what could be understood as a pioneer form of development aid. I analyze the factors influencing the origin of these associations, their decision to construct schools in their villages, and their impact on educational attainment and further migration.

The remainder of the paper is structured as follows. Section 2 provides some historical background. Sections 3 and 4 describe the data and empirical strategy. Section 5 and 6 discuss the results and mechanisms. Finally, Section 7 concludes by drawing some policy implications.

2   Historical background

2.1   Galician mass emigration

Galician mass emigration in perspective. The Galician diaspora is one of the largest emigration episodes in the twentieth century. Between 1900 and 1930, over 1.1 million Galicians left this Spanish region, which represents 58% of its initial 1900 population (Vázquez-González, 1999, p. 375).[9] As shown in Table 1, Galicia had higher gross emigration rates than any other European country at the time. Return migration was relatively high, close to 70% during 1900-1930. Based on these figures, Galicia accounts for around 45% of Spanish net migration to Latin America (Vázquez-González, 1999, p.379). Galician mass migration was concentrated in 1900-1930, stopping abruptly in the early 1930s after Latin American countries underwent severe economic crises and introduced migratory restrictions. The yearly flows and returns are displayed in Figure A2.

Causes of migration. The first emigration waves to Latin America took place in 1840-1860 and it is

---

[9]This estimate is a lower bound since official statistics do not capture clandestine migration that could exceed 20% (Sánchez-Alonso, 1995). For the 1835-1930 period, Vázquez-González (1999, p. 374-75) estimates between 1.3 and 1.6 million departures while Eiras Roel (1991) a figure of 1.7 million (for 1840-1930). In terms of return migration, the official statistics document 600 thousands in 1900-1930 although the data on returns may be even more underestimated than that of departures.

argued to have been caused by agricultural and rural textile industry crises (Bertrand, 1997; Eiras Roel, 1991; Hernandez, 1999). As shown in Figure 1, rainfall levels were extraordinary high and volatile in those decades, what could have been detrimental for the cereal harvests and the flax used in textile. The effect of these shocks would have been accentuated by an archaic agriculture (Álvarez Campos, 1993; Farías, 2010) and a high demographic pressure (Eiras Roel, 1991, Vázquez-González, 1993). While emigration kept increasing steadily during the following decades, it remained relatively low. In the twentieth century, however, it reached unprecedented levels fostered by the rapid economic growth in Latin America, transportation innovations[10] and the contribution of migrant networks (Álvarez Campos, 1993; Farias, 2010).[11] Importantly, Galicia already had a tradition of seasonal migration to the North of Portugal and Castille during the XVIII and XIX centuries (Farías, 2010).

Destinations. The main destinations were Rio de la Plata (58%), Cuba (32%), and to a lower extent Brazil (9%). Due to the strong influence of migrant networks, most migrants from a given area tended to go to the same destination.[12] Yet, as illustrated by Figure 3, there was a large geographical heterogeneity in the relative importance of each country. Section B in the Appendix provides more details about migrant destinations.

Migrants' profile. Galician migrants were mostly men (over 70%), young (50% aged 16 to 30) and rural farmers or craftmen (90%) (Vázque-González, 1999).[13] Migration was a family strategy under which one or more members would go to Latin America temporarily to complement household income and to reduce the burden at home. Once at destination, they located almost exclusively in large urban centers and worked in services (e.g., small stores, catering, retail, storage, etc.). Migrants usually starting working for a relative or acquaintance until having enough savings to set their own business and eventually return after a few years (Álvarez-Campos, 1993; Vázquez-González, 1999).

---

[10]In the 1870s, sailing from Galicia to Cuba and Argentine took around one month and half and two months respectively. The spread of steamships in the 1880s reduced drastically travel times, making it possible to reach Cuba in 19 days and Argentine in 25 days (Vázquez-González, 1993).

[11]Other factors stressed in the literature are the difficult access to land due to the "foros" regime and the heritage system (Betrand, 1997), the high tax burden (Eiras Roel, 1989), and military desertion (Álvarez Campos, 1993). See the Appendix for a brief summary of the economic history of Galicia.

[12]Vázquez-González (1999, pp.882-886) and Seixas-Núñez and Vazquez (2004) show that this pattern is also present across parishes within the same municipality.

[13]These figures are a probably a lower bound since clandestine migration was characterized by young men from rural areas (Vázquez-González, 1999, p. 383).

Migrant associations. In the mid-nineteenth century Spanish migrants created associations of mutual insurance providing health care to their members.[14] These were followed by regional associations in the 1870s and local associations in the twentieth century.[15] Between 1900 and 1930 hundreds of Galician local associations were created thanks to the large inflow of migrants and their concentration in the same cities in Latin America.[16] The grounds for local associations were often an informal gathering, for example, to celebrate a feast in honor of the village's saint (Fernández-Santiago, 2008). Through these associations migrants raised funds to invest in public goods in their hometowns such as roads, schools, etc. While the membership of local associations started to decline in the 1930s (until their extinction in the following decades), regional Galician centers have persisted until nowadays. Most migrant associations had an educational goal and among their main objectives was the construction of schools in their hometowns. Over 300 hundred schools were founded by these migrant associations in the first half of the twentieth century, representing around 7% of the stock of public schools at the time. Figure A22 and Figure A23 in the Appendix show the foundation of associations and schools over time and their location.

2.2 Education in Galicia

Until the first decades of the XX, the education system in Galicia presented a large deficit of material and human resources, low attendance rates, poor educational outcomes and large gender inequalities (De Gabriel, 2006). While the education laws required a minimum number of schools depending on municipalities' population, the highly dispersed settlements within Galicia municipalities led to the closest school often being located several kilometers away and to a ratio of schools per children lower than the national average.[17] The difficult access to schools was aggravated by the abrupt geography and poor transportation infrastructure. The conditions in most establishments were penurious; muddy ground, lack of light and ventilation, shared materials,

---

[14]Some examples are "Sociedad de Beneficiencia de Naturales de Cataluña", founded in La Habana (Cuba) in 1840, "Sociedad Española de Beneficiencia" and "Asociación Española de Socorros Mútuos" founded in Buenos Aires (Argentine) in 1852 and 1857 respectively. The first Galician association of this kind appeared in 1871 in La Habana: "Sociedad de Beneficiencia de Naturales de Galicia" (Fernández-Santiago, 2008).

[15]The Centros Gallegos were founded simultaneously in La Habana, Buenos Aires and Montevideo in 1879. In Buenos Aires there were also associations from other regions: Centro Aragonés (1885), Centre Catalá (1886), Centro Balear (1905) and Centro Asturiano (1913). The first local associations of Galician migrants appeared in 1904 ("Alianza Arensana de Instrucción" in Cuba and "La Concordia" in Argentine).

[16]Vázquez-González (1999) documents 665 of them, among which over 400 had an educational goal.

[17]Moyano law (1853) required at least one school in municipalities with more than 500 inhabitants and two schools if population exceeded 2,000. While in 1910 Galician municipalities had on average around 6,000 inhabitants (compared to 2,100 in Spain), people often lived in parishes of less than 500 people (Martínez-Dominguez, 2000, p. 49). The average number of parishes in Galician municipalities is 12.

etc. Teachers lacked qualification and in many municipalities this role was done by a priest.[18] The curriculum consisted in Catholic doctrine (which involved learning how to read), writing, basic arithmetic and domestic tasks in the case of girls (Martínez-Domínguez, 2000, p. 50).

The poor quality offered, the obstacles to access schools, and the low value parents put in education made school enrolment very low and attendance irregular, peaking in winter when children were less demanded for agricultural tasks (De Gabriel, 1970). The lower enrolment of girls compared to boys has been argued to be the result of parents prioritizing boys' education given that they would be the future household heads and because it could be more profitable in the prospects of migration (Vázquez-González, 2002; Martínez-Domínguez, 2002).[19]

3 Data

This paper draws on newly digitized data from multiple historical sources. A summary of all the sources can be found in Table B1 and Table B2 in the Appendix.

Population censuses. The main data comes from the seven population censuses of Spain between 1860 and 1930.[20] I digitized all the information available for the 321 Galician municipalities; the number of present, missing, transient and literate individuals by sex. Household members were classified as present or missing depending on whether they had been in the municipality the two nights prior to the census day. I use the share of missing men in a municipality as proxy for migration and discuss its validity in the next section. Transients are individuals found in a different municipality than the one where they regularly reside in, typically representing less than one percent of the population. Household heads reported literacy information for all present members. I compute literacy rates as the share of present individuals knowing how to read and write.

Missing men as proxy for migration. Due to the absence of migration data at the local level, I use the share of missing men as proxy for the stock of emigrants. This measure presents the advantage of being available for all municipalities and census years in a comparable manner. Moreover, it

---

[18]In 1855, around 51% of teachers in public schools in Galicia had a degree compared to 81% in Spain (Martínez-Domínguez, 2002). Moreover, until the 1870s most schools had only one teacher and a single group of students with different ages.

[19]Note that the supply of school played no role in gender disparities as most establishments were mixed. Even once attending school, girls received fewer hours of instruction in literacy and numeracy compared to boys (Martínez-Domínguez, 2002).

[20]The Spanish population censuses were conducted in 1860, 1877, 1887, 1900, 1910, 1920 and 1930. Figure A3 in the Appendix provides an illustration of the original documents.

captures clandestine migration that at the time could exceed 20% (Sánchez-Alonso, 1995). Note that the census was conducted the 31st of December precisely to minimize the number of missing and transient individuals as most people were expected be home those days.[21] I conduct a number of descriptive analysis to examine the validity of this measure. First, the figures of missing men and net migration in Galicia are almost identical and follow the exact same trend over the 1877-1930 period (Figure 5). Similarly, the share of missing men and gross emigration rates are strongly correlated across the Galician provinces in different years (Figure 6), but also across the 47 Spanish provinces (Figure A4) and even the Italian provinces (Figure A5). Second, the transient population in Galicia remained constant and below one percent over the period, what supports the idea that missing men were not in neighboring municipalities or elsewhere in Spain (Figure A6).[22] Galicians where the least likely to live in another Spanish region (3% compare to an average of 10%). Additionally, the number of drafted men in Galicia was fairly stable in absolute and relative terms suggesting that the draft played no role driving the trends in missing men (Figure A7).[23] Note that I do not rely on the share of missing women for a number of reasons. First, as shown in Figure A4, when emigration flows are low the figures on missing residents and sea departures are not strongly correlated. Between 1900 and 1930 the share of missing women ranged only between 2% and 3%, which suggests it may not accurately capture female departures but rather other transitory movements out of the household. Moreover, there were tighter restrictions to women's migration at the time, which made them more likely to emigrate to join family members at destination and hence with a more permanent nature (Sánchez-Alonso, 1991). As a result, female migrants were more likely to not be reported as missing in their previous household unit. I will show nonetheless that the results of the analyses are very similar using the share of missing residents instead of the share of missing men.

Sex ratios as proxy for migrant networks. Since the 1860 population census does not contain information on missing residents, I rely on the adult sex ratio as proxy for the intensity of the first migration waves to Latin America (1840-1860). The pioneers left Galicia mostly as a response to

---

[21] *"La fecha fijada por la ley para llevar a efecto la inscripción de habitantes es la más apropiada para obetener la población en su mayor estado de reposo, o sea para reducir a una cifra mínima el número de ausentes y el de transeuntes"* (Spanish population census 1920, p. 68).

[22] It is unlikely that missing men reflects internal migration since it was mostly concentrated between the months of May and September (Palmás, 1984).

[23] Given that there were no consequences of having missing household members, household heads should have no incentives to miss-report. Individuals were warned that they would be subject to a fine if they provided inaccurate information.(XX)

the extreme weather conditions of those decades and their impact on agriculture. Figure 2 shows that there is a strong correlation between the adult sex ratio and the intensity of rainfall variability in 1830-1860.

Economic conditions at destination as pull factor. Galician migrants were attracted by the economic conditions in Latin America. I use data from the last release of the Maddison Project Database (2018) to construct a measure of GDP per capita growth in the average destination. To infer the main destinations across Galician municipalities I digitized information compiled by Vázquez- González (1999) of over 17 thousand people emigrating during 1900-1930.[24] This information comes from a variety of sources: embarkation lists, notarial documents and desertion records. As a robustness, I use official data of sea departures at the province level in 1885-1895, the earliest period this data is available. Section B in the Appendix describes how the pull factor measures were constructed. Figure C1 shows the variation in economic growth at destination across municipalities and over time.

Migrant associations and schools. I constructed a database of over 400 Galician migrant associations in Latin America with information on the country and year of their foundation, the Galician municipality to which they relate, their intervention focus, the schools they constructed in Galicia and the opening year of such schools. The sources used are detailed in Table B1 in the Appendix.

Contemporary data. I rely on two alternative data sources to obtain measures of educational attainment: the population censuses (1981-2011) and the regional household surveys (2007-2018). To trace internal mobility, I use administrative data from the population registers ("padrón" of 1998 and 2003). The surveys "Opinión Pública y Política Fiscal" by the research center CIS provide me with information on attitudes towards education, value of effort and preferences over redistribution (2011-2018). Additionally, I use data on schools and enrollment at the municipality level from the official statistics of 1975 and 2018 and information about the universe of local associations (around 32 thousand in 2018) and public libraries.

Alternative proxy for migrant networks. I use the walking distance to the closest transatlantic

---

[24]The data prior to 1900 is much more limited but there seems to be a high degree of persistence in the main destinations. For instance, the correlation between the share of migrants going to Rio de la Plata in 1900-1930 and in 1850-1860 across districts is 0.54.

port as an alternative proxy for the size of migrant networks. Due to the poor transportation infrastructure of Galicia, being close to a port was crucial for the diffusion of information and for the establishments of stronger migrant networks.[25] For each municipality, I calculate the shortest distance to a port following the road network of the nineteenth century based on an official historical map that I geo-referenced (Figure E1). Section D of the Appendix provides further details.

Rainfall shocks at origin as push factor. I use data from the European Gridded Seasonal Precipitation Reconstructions (Pauling et al., 2006) to construct measures of weather shocks. The fine resolution of the data-set ($0.5° \ x \ 0.5°$) yields 24 cells for Galicia. I focus on rainfall during the Fall season given that maize was the main crop cultivated in the region and it was harvested in that period. For every decade, I compute the standard deviation of rainfall as shown in Figure D1.[26]

Supplementary historical data. I exploit supplementary data coming from multiple sources described in Table B1 in the Appendix. These include newly digitized data at the municipality level on wealth and taxation (Madoz, 1845), schools and enrollment (1875, 1908), and statistics of livestock across districts (1865, 1897, 1917).

4 Empirical strategy

The objective of this paper is to examine the impact of mass emigration on human capital accumulation at origin over a century. Fist, I use a decadal panel of Galician municipalities in 1900-1930 to understand the determinants of emigration (Section 5.2) and analyze its short and medium-run impact on literacy rates (Section 5.3). I then look at the long-run effects of historical emigration (1900-1930) on different education-related outcomes using a cross-section of all municipalities (Section 5.4). Finally, I explore different mechanisms exploiting additional historical data as well as contemporary administrative and survey data across individuals, cohorts and municipalities (Section 6). Below I describe the different empirical strategies and the main identification challenges in each of them.

---

[25]Vázquez-González (1999) documents the areas of activity of several migration hooks in the 1850s. These agents operated in municipalities surrounding the ports of Coruña, Vigo and Vilargacía and expanded their activities inland over time.

[26]This data has been recently used as push factor in other works such as Sequeira, Nunn, Qian, 2020. Section C in the Appendix provides more details.

4.1 Short and medium-run effects [Panel analysis]

Estimating equations. To examine the dynamic relationship between changes in missing men and literacy rates, I exploit a panel of 321 municipalities during four periods (1900, 1910, 1920, 1930). I refer to the short-run impact as the contemporary relationship and to medium-run impact as the one observed one decade later. The main estimating equations are:

Short-run effect:

$$Literacy_{it} = \alpha_i + \alpha_t + \phi ShMissing_{it} + X_{ijt-1}\Gamma + \epsilon_{it} \quad (1)$$

Medium-run effect:

$$Literacy_{it} = \alpha_i + \alpha_t + \psi ShMissing_{it-1} + X_{ijt-2}\Theta + v_{it} \quad (2)$$

where $Literacy_{it}$ is the share of people knowing how to read and write in municipality i and year t, θ_i and θ_t represent municipality and year fixed effects, $ShMissing_{it}$ is the share of missing men in year t and ϵ_{it} stands for robust standards errors clustered at the municipality level. The vector X_{ijt-1} represents lagged controls (i.e., population, literacy rates and GDP per capita in the province). $ShMissing_{it-1}$ represents the share of missing men mesaured one decade before.

Estimating equations (1) or (2) by OLS may give rise to three concerns. First, measurement error in the share of missing men and literacy rates would lead to attenuation bias and less precise estimates. Note that the share of missing men is an imperfect proxy of the stock of emigrants and other scholars have identified inconsistencies in literacy information in the census (De Gabriel, 2006). Second, there are potential confounders of migration that could bias the results in different ways (e.g., a negative income shock in a municipality could affect both investments in education and migrant flows). Third, the analysis could suffer from reserve causation if a raise in literacy rates affected the diffusion of information about the prospects abroad (Sánchez-Alonso, 1995) or if it deterred migration by revitalizing the local economy. To deal with these issues, I implement an instrumental variables strategy exploiting two sources of plausibly exogenous variation.

Instrumental variables. I instrument the share of missing men with a pull factor (average GDP per capita growth at destination) and its interaction with a proxy for migrant networks (the adult sex ratio in 1860). The intuition is that faster economic growth at destination would attract more migrants, but its impact could be amplified by migrant networks. Recent empirical work has

stressed the prominent role of networks during the Age of Mass Migration (Spitzer & Zimran, 2018; Brum, 2019) but also in the Galician case (Vázquez-González, 1999). While the proxy for the migrant network is constant over time, the pull factor varies according to the economic growth at the average destination. The strategy exploits thus variation provided by these two forces at the local level. The first stage equation is:

$$ShMissing_{it} = \theta_i + \theta_t + \delta\, Pull_{it} + \rho\, Pull_{it} \times Network_i + X_{ijt-1}\Pi + u_{it} \quad (3)$$

where i indexes municipalities and t years. The variable $Pull_{it}$ represents the GDP per capita growth at the average destination country, $Network_i$ represents the adult sex ratio in 1860, X_{ijt-1} a set of lagged controls and u_{it} standard errors clustered at the municipality level. The pull factor is constructed using information on the destination of over 17 thousand Galician migrants between 1900 and 1930. Formally:

$$Pull_{it} = \sum_c w_{ic} \times \Delta y_{ct} \quad (4)$$

where w_{ic} is the share of migrants from municipality i going to country c during 1900-30 and Δy_{ct} the GDP per capita growth of country c in the previous decade. As shown in Figure C1, there is a lot of geographical and time variation in this measure because municipalities were connected to different countries (Figure 3) and their economic growth was not correlated (Figure 4).

Identifying assumption. The identification requires that economic growth at destination is not correlated with any time-varying factor that could affect literacy rates differently depending on the initial sex ratio in 1860. Note that the identification does not require the sex ratio in 1860 to be exogenous as the effect of any time-invariant variable would be captured by the municipality fixed effects. The strategy would be flawed for instance if (i) economic growth at destination and origin were correlated and (ii) if the adult sex ratio in 1860 was correlated with economic activity.[27] None of these hypotheses seems to be supported by the data. Figure A9 shows that the evolution of GDPpc at destination is uncorrelated with changes in economic activity at origin. This result is not surprising given that Latin America represents less than 10% of Spanish trade in 1900-1930 and that Galicia was not particularly open to international trade in that period (see **??**). Goldsmith-Pinkham, Sorking and Swift (2019) highlight that the exogeneity of the time-invariant

---

[27]This would imply that the instrument predicts higher emigration (when high economic growth in Latin America and having a lower sex ratio) as well as an changes in income levels at origin that could in turn affect literacy rates.

component is crucial in a shift-share like instrument. The fact that the first migration wave (1840-1860) was mainly caused by extreme rainfall gives support to the idea that the adult sex ratio in 1860 may be exogenous, or at least conditional on covariates. I test formally the relationship between the adult sex ratio in 1860 and several pre-migration variables (1850-1860) showing that it is not correlated with educational inputs or the evolution of population or literacy rates between 1860 and 1900 (Figure A10). Recent work by Jaeger, Ruist and Stuthler (2019) argue that shift-share instruments may confound the shor tand long-run effects of immigration. This concern is less likely to apply in my analysis given the decade nature of my data. Similarly, Christian and Barret (2017) caution about the use of instruments with a time varying factor in an interaction as there could serial correlation at the aggregate level. My strategy does not suffer from this concern given that literacy rates experience an upward trend over the period while missing men display an inverted-u pattern. Note that I will perform a series of robustness checks to rule out that my results are driven by confounders.

4.2 Cross-sectional long-run analysis

Estimating equation. To examine the long-run effects of historical emigration on contemporary outcomes I rely on a cross-section of all Galician municipalities. The main explanatory variable is the average share of missing men in 1900-1930 that I instrument with the adult sex ratio in 1860, a proxy for pioneer migration. The estimating equation is:

2nd Stage:

$$E_i = \alpha + \beta \, \overline{ShMissing_i} + Z_i \Omega + \omega_i \quad (5)$$

1st Stage:

$$\overline{ShMissing_i} = \lambda + \gamma \, Network_i + Z_i \Phi + e_i \quad (6)$$

where E_i represents different educational outcomes, Z_i historical / contemporary controls, and $Network_i$ is the adult sex ratio in 1860. The standard errors ω_i and e_i are clustered at the district level (n=47).

Identifying assumption. The identification strategy requires that the adult sex ratio in 1860 is uncorrelated with other factors affecting human capital accumulation over the twentieth century. Although this assumption cannot be tested directly the balance test performed in Table A4 shows that the adult sex ratio in 1860 is not correlated with relevant covariates such as wealth per capita

(1855), local taxes per capita (1855), the supply of schools (1875), school enrollment rates (1875), changes in population or in literacy (1860-1900). There is a small (yet significant correlation) between the sex ratio and literacy rates in 1860, however it becomes insignificant once province FE are introduced. Controlling for different historical factors barely affects the estimates of the first stage nor the second stage. In the regressions I will control for population size, literacy rates, wealth per capita and local taxes in 1855-1860 although this inclusion makes no difference in the results.

4.3 Cohort analysis

Estimating equation. To evaluate the impact of migrant associations and the schools they built in their hometowns I perform a cohort analysis. The identification is very similar to Duflo (2001) with the addition of considering several cohorts given that associations and schools opened at different years in 1900-1940. I exploit data from the 1981 population census with information on the number of individuals by birth cohort (5-year groups) and educational attainment across municipalities. I restrict the sample to individuals born between 1902 and 1927 (aged 54 to 79 in 1981). The estimating equation is:

$$y_{ci} = \theta_c + \theta_i + \alpha \ Association_{ci} + \beta \ School_{ci} + v_{ci} \quad (7)$$

where y_{ci} represents the outcome of individuals born in municipality i (living there in 1981) from birth cohort c , θ_c and θ_i represent cohort and municipality fixed effects respectively, $Assoc_{ci}$ and $School_{ci}$ represent the number of associations / schools openened in municipality i when the birth cohort c was below 10 years old or younger.[28]

To explore the impact of migrant associations on net migration, I rely on the ratio between size of a cohort c and that of cohort $c - 1$. Given that individuals are observed when aged 54 to 79, all emigrants coming back to their home municipalities would be included. To measure educational attainment I use the share of individuals who have completed at least primary education (e.g., 50% of men and 27% of women).

Identifying assumption. A causal interpretation of equation (7) relies on the assumption that there should be no time-varying local factors correlated with the opening of associations or schools

---

[28]This measure is takes into account the relative exposure of individuals within a cohort window. For instance, consider the cohort aged 6-10 in 1920; if the first migrant school opens in 1920, then all the individuals in that cohort group would be exposed and $School_{1910-14,i} = 1$, if the school opened in 1921, individuals aged 10 in 1920 would be untreated while younger ones fully treated, meaning $School_{1910-14,i} = 4/5$.

and affecting investments on children. In particular, there should be no changes in mortality or migration for the cohorts affected. Moreover, the timing of the opening of associations and schools should be uncorrelated with the expected impact, that is assumed to be constant across municipalities. To examine the validity of these assumptions I will estimate equation (2) treating different combinations of cohorts as exposed and I will implement a more flexible dynamic event study (Abraham and Sun, 2018). Note that an important caveat is that migration flows may influence the delay between the foundation of an association and the opening of a school, however, there is no simple way of tackling this issue as migration is a necessary condition for migrant's investments.

5 Results

5.1 Descriptive evidence: Evolution of literacy rates and the share of missing men

This section describes briefly the trends in missing men and literacy rates between 1860 and 1940. During the second half of nineteenth century, emigration affected mostly western coastal municipalities, which saw their share of missing men reach between 10% to 20% by 1900. During the first two decades of the twentieth century, the share of missing men raised significantly in the whole region, especially inland and in northern coastal municipalities. By 1920, the typical Galician municipality had over 18% of men were missing the day of the census. The decade of the 1920s saw a slight decrease in the share of missing men in most municipalities and with the stop of migrant flows during the 1930s the the share of missing men came back to levels of 1900.[29]

Literacy rates experienced a rapid increase between 1860 and 1940, especially after 1910 (see Figure 9). The initial gap between men and women literacy rates narrowed over time and almost disappeared by 1940. During the nineteenth century, literacy rates were not significantly different across municipalities that would experience higher/lower emigration later on. Yet, after 1900, a gap starts to emerge with high emigration areas seeing a faster increase in literacy rates. Figure 10 shows the evolution of the gap in literacy rates between municipalities above and below the average share of missing men (1900-1930). The gap peaks in 1920 at 2.5 percentage points and falls afterward. The similar evolution of the share of missing men and the gap in literacy rates between areas differently exposed motivates a deeper investigation of the potential relationship between emigration and human capital accumulation.

---

[29]Figure 8 in the Appendix describes the spatial distribution of missing men in each census year.

5.2 Explaining the share of missing men, 1900-1930

Before examining the short and medium-run effects of emigration on literacy rates, in this section I discuss the main determinants of the share of missing men.

PANEL ANALYSIS. Table 2 shows the results of estimating equation (3), a panel regression of the share of missing men in year t on different lagged covariates. All the columns include municipality and year fixed effects capturing the influence of time-invariant local factors and common trends across all municipalities. The table displays beta coefficients to make interpretation easier. In column (1), the share of missing men is regressed on the main instruments, namely economic growth in the average destination and its interaction with the adult sex ratio in 1860. At the average level of the adult sex ratio (i.e., setting the interaction effect to zero), an increase in GDP per capita growth at destination in one standard deviation is associated with a raise in the share of missing men in 7% of a standard deviation (e.g., 1 percentage point). Increasing the adult sex ratio in one standard deviation amplifies this effect by more than half (i.e., an additional 4 percent) leading to an increase in the share of missing men of 1.5 percentage points. This effect is non-negligible, especially given that the fixed effects already capture a large part of the variation in the share of missing men. The coefficients are very robust to the inclusion of other lagged controls even the lagged share of missing men (columns 2 to 4). The last column displays the result including all covariates. There is a high degree of persistence in missing men. On average, one standard deviation increase in the lagged share of missing men is associated with a 24% of a standard deviation increase in the contemporary share. Rainfall shocks are negatively related with migration, suggesting that migrants may have faced credit constraints. Province-level factors such as population and literacy rates are also strong drivers of aggregate changes in missing men.

CROSS-SECTIONAL ANALYSIS. Table 3 shows the results of estimating equation (6), a cross-section regression of the average share of missing men (1900-1930) on the adult sex ratio (1860) and different pre-migration covariates. The table displays beta coefficients to make interpretation easier. Column (1) shows that the instrument, the adult sex ratio in 1860, is a key predictor of the share of missing men explaining close to 20% of its variation. Columns (2) to (4) add geographic, socio-demographic and economic variables. The coefficient of the adult sex ratio remains very robust across the different specifications. On average, an increase in the adult sex ratio (1860) in one standard deviation is associated with a raise in the average share of missing men (1900-1930)

in 30% of a standard deviation (i.e., a 2.8-3.5 percentage points relative to a mean of 14 percent). Column (5) includes all the covariates at once. The results show that bigger municipalities and those far away from a port had significantly lower shares of missing men during 1900-1930. Distance to the port can be interpreted as reflecting transportation costs. Regarding initial literacy rates (1860), there is a positive correlation in the case of men but a negative one in the case of women. These findings suggest that men literacy rates facilitated migration, either by allowing high educated individuals to migrate or by contributing to the diffusion of information (as hypothesized by Sánchez-Alonso, 2000). Land inequality and local taxes are negatively associated with the share of missing men (although the latter is not statistically significant), supporting the idea that migrants may have been credit constrained. Regarding transportation infrastructure, access to a train network and to the sea are both strongly and negatively associated with the share of missing men but imprecisely estimated, potentially given that the regression already controls for distance to the port and to province capitals. Although the role of population density has been stressed in other works (Eiras Roel, 1989; Vázquez-González, 1993), I do not find any evidence that population pressure in 1860 influenced the extent of missing men later on.

ROBUSTNESS CHECKS: The significance of the instrument and the point estimates remain very robust across different specifications in the panel and cross-sectional analysis. Table A3 shows the relationship between each instrument and the share of missing men in different years.

5.3   Short and medium-run effects of emigration on literacy rates, 1900-1930

In this section, I describe the short and medium-run effects of mass emigration on literacy rates. I rely on a decadal panel of municipalitites between 1900-1930 and instrument the share of missing men (in a given year t) by the economic growth at migrants' destination in the precedent decade (between t-10 and t) and its interaction with the adult sex ratio in 1860.

MAIN RESULT GRAPHICALLY. I first plot the coefficients from estimating variants of equation (1) by 2SLS with different lags in the share of missing men or literacy rates. As shown in Figure 11, the estimates of t-10 and t-20 suggest that the share of missing men in a given year has not a strong impact on the literacy rates one or two decades before. This exercise serves as a placebo test of the instrumental variables strategy. The estimate of t shows that the contemporary effect of missing men on literacy rates is large and negative implying a reduction in literacy rates in the short run.

In the medium run the impact of missing men is positive and significant implying a net gain in literacy rates one decade later (t+10) which persists two decades after the migrant shock (t+20).

MAIN RESULT IN DETAIL. Table 4 summarizes the results on the short and medium-run impact of missing men on literacy rates (i.e., OLS and 2SLS estimates of equations 1 and 2). The coefficient in column (1) shows that taking into account municipality and year fixed effects there is no correlation between the share of missing men and literacy rates. Using the share of missing men in the previous census (10 years before) yields a small positive and significant correlation as shown in column (2). The 2SLS estimates suggest that on average one standard deviation increase in the share of missing men in year t decreases literacy rates in t by 51% of a standard deviation but increase literacy rates 10 year later by around 35% of a standard deviation. The 95% confident intervals around this estimate imply that a migration shock raising the share of missing men from 11% to 20% could increase literacy rates one decade later by around 0-6 percentage points (i.e., an increase of 0%-17%).

Note that there is a substantial gap between 2SLS and OLS estimates, although the coefficients are not statistically different from each other. The short-run OLS estimate is upward biased, which could be the result of positive shocks affecting simultaneously literacy and emigration. There is recent evidence documenting that economic growth can indeed foster emigration (Clemens and Mendola, 2020). Similarly, there could be reverse causation to the extent that increases in literacy rates facilitate the diffusion of information about the option to emigrate. Sanzhez-Alonso (1995) examines this hypothesis using data across Spanish provinces at the beginning of the twentieth century and finds that higher literacy rates are associated with more emigration. Notice that these two sources of bias would tend to disappear in the medium run to the extent that other areas experience similar positive shocks. Interestingly, the medium-run OLS estimate is downward biased. One potential explanation is that, despite an eventual short-run positive shock, areas experiencing higher emigration are generally on less favorable development paths with gaps that tend to widen over longer horizons. It is crucial to point out that both the short-run and medium-run OLS estimates are biased toward zero. This could be the result of measurement error in the the share of missing men. As discussed in the Data section, the share of missing men is a valid but imperfect proxy for emigration and the instrumental variable strategy would tackle this attenuation bias.

GENDER DIFFERENCES. The impact of emigration on literacy rates differs by gender as shown in

Figure 12. While there is a short-run negative effect for both men and women, the medium-run gains are only present for men. In the short-run, an increase in the share of missing men in one standard deviation (i.e., 9 percentage points) leads to a decrease in male and female literacy rates of 33 and 55% of a standard deviation respectively. One decade later, male literacy rates are 66% of a standard deviation higher (i.e., around 6 percentage points) while no significant effect is found for women. The fact that the coefficient for women is close to zero suggests that the short-run negative effects dissipate but there are no gain arising in the medium run. Although these findings are consistent with the idea that families may invest more in the education of individuals who have higher returns to schooling (Vanden Eynde, 2016; Shrestha, 2017), in the Mechanism section I provide suggestive evidence that this is not the case, as emigration contributed to equally higher investments on both boys and girls' education. The more plausible explanation for the gender disparities in the medium run might be differences in the degree of return migration and selection into return.

HETEROGENEOUS RESULTS. I explore the potential heterogeneity of the short and medium-run effects depending on migrants' destinations (1900-1930), population size and initial literacy rates before the mass emigration. The results are summarized in Figure A13. First, I divide the sample between municipalities with a share of migrants to Rio de la Plata above and below the median. The results suggesting that at least in the medium run, migrants' destination did not influence the effects on literacy rates. I then divide the sample between municipalities with a population above (High population) and below (Low population) the median in 1860. In the case of men, the negative short-run effects are entirely driven by areas that had relatively higher population. In the medium run, however, the gains in literacy rates are of similar magnitude everywhere. In the case of women, the estimates tend to be larger in municipalities with higher population in 1860 but the coefficients are not statistically different from one another. Finally, I divide the sample between municipalities with literacy rates above (High literacy) and below (Low literacy) the median in 1860 (gender-specific). For men, the short-run and medium-run effect are larger and only significant in municipalities with relatively higher male literacy rates in 1860. The coefficients however are not statistically different. In the case of women, the differences are smaller, potentially because in 1860 female literacy rates were extremely low in all municipalities.

ROBUSTNESS CHECKS. I perform a number of tests to assess the robustness of my main results

and address potential threats to the identification:

Alternative instruments. The main results remain very similar by using alternative pull factors constructed using district weights for the period 1900-1930 or province weights for the years 1885-1895 as shown in Figure A14. Being able to replicate the findings using weights before 1900 shows that the selection of migrants to certain countries is not confounding the results. The estimates are also very similar when using the walking distance to the port or directly as network instrument .

International trade. One potential concern is that there may be a relationship between economic growth at destination and Galician trade with that country. Controlling for trade with destination countries, its interaction with the walking distance to the port (columns 1 and 2 of Table A8), or excluding municipalities closer to the ports (columns 1 and 2 of Figure A15) has no effect on the main results. Note that Latin America represented a small fraction of Spanish exports and imports, and Galicia was not particularly open to international trade in 1900-1930.

Remittances and economic shocks. Given that I exploit an instrument that captures the size of the network, one may worry that the effects of remittances get amplified and therefore influence the results on literacy rates. To partially deal with this concern, I add different measures of economic activity at origin (e.g., GDP per capita in the province, GDP per capita growth in the province, interaction between GDP per capita in the province interacted with distance to the closest port, etc.). I also introduce measures of rainfall shocks at origin that by changing the level of remittances (or their use) could alter the impact of emigration on literacy rates. Overall, including these variables has no effect on the estimated coefficients as shown in Table A8.

Extensive controls. To make sure that none of the potential (observable) confounders of the adult sex ratio (1860) is mediating the results, I introduce the pre-migration covariates (1855-1860) interacted with time dummies. The results show that the instrument is unaffected by the inclusion of these variables and that the main findings remain robust.

Sample restrictions. I analyze whether the estimates are sensitive to outliers by excluding municipalities with extreme values of population, missing men, literacy rates or changes in literacy rates (see Figure A15). I also exclude provinces one by one to check if the effects differ geographically. The coefficients remain very similar across the different specifications.

Missing women. I replicate the analysis replacing the share of missing men with the share of missing population (i.e., men and women absent the day of the census) and the share missing women alone. The results are displayed in table Table A9. As shown in columns (1) to (6), using the share of missing population produces very similar estimates of the short and medium-run

effects. The strength of the first stage is however significantly lower due to the noise introduced by incorporating missing women into the explanatory variable. Columns (7) and (8) look at the impact of the share of missing women on women's literacy rates but in this case the instruments are not fitted for the specification, producing an F-stat of 1.28 and 4.35 respectively.

5.4 Long-run effects of emigration on human capital, 1930s-2000s

In this section, I describe the long-run effects of mass emigration on different measures of human capital between 1930 and 2011. I estimate equation (5) with a cross-section of all Galician municipalities instrumenting the average share of missing men (1900-1930) with the adult sex ratio in 1860 (i.e., a proxy for pioneer emigration).

MAIN RESULT. Table 5 summarizes the results on the long-run effects of emigration showing that it contributed to gains in average levels of human capital that have persisted until today. Focusing on 2011, the year of the last census, the estimate from column (4) suggests that one standard deviation increase in the average share of missing men (1900-1930) led to a higher share of individuals with completed secondary education by 46% of a standard deviation (i.e., 6 percentage points relative to a mean of 59 percent). Similar to the results on the medium-run effect of missing men in the panel analysis, the OLS coefficients of the long-run effects are downward biased and less precisely estimated than 2SLS. This result supports the idea that areas that experienced more emigration (on aggregate) had characteristics that lower their economic potential and limited human capital accumulation over the long run. As discussed in section 5.2, the average share of missing men (1900-1930) is positively associated with more isolated areas (e.g., further away from a province capital, less likely to have a train connection, higher average altitude, etc.) and prone to greater rainfall variability which could have had a negative and persistent influence on educational investments. [30]

I further explore the impact on different levels of education for the cohorts born between 1930 and 1980 using data from the 2001 population census. As shown in Table A10, mass emigration seem to have affected human capital accumulation from the bottom of the distribution up. First, increasing literacy rates in the period while emigration was taking place (1900-1930), then raising primary completion rates for cohorts born in the 1930s and thereafter, secondary completion rates

---

[30]Controlling for these factors in the OLS regression reduces the downward bias slightly but the coefficients remain quite small, pointing to the relevance of attenuation bias.

for those born in the 1930s and 1940s, and finally tertiary education rates for those born in the 1940s and 1950s. Interestingly, there is a zero impact of missing men on tertiary completion for individuals born in the 1960s and a large negative effect among those born in the 1970s. These findings may suggest that emigration could have fostered these generations to attend college and, as a consequence, to temporarily leave their municipalities of origin and decrease the overall share of individuals with completed tertiary education. Overall, my results contrast remarkably with the findings of Testa (2020), which reports a negative long-run effect of forced migration on human capital at origin six decades later. Yet, as mentioned previously, episodes of forced migration and voluntary migration are not comparable due to the particularities of the former.

GENDER DIFFERENCES. Figure 13 shows that the long-run effects of emigration differ by gender but converge over time. While in 1981, there is no evidence of a significant impact for women, the share of missing men (1900-1930) positively affected secondary completion among men. These results are in line with the findings for the historical period (1900-1930), where men experienced gains in literacy rates in the medium run while women did not. In 1991, the coefficient for women is positive and marginally significant and in 2001 and 2011 becomes much larger in size and highly significant. Importantly, if one restricts the sample to women with less than secondary (illiterate, no studies completed, primary completed), there is a positive effect in 1981 for the share who had completed at least primary suggesting again that emigration raised educational attainment bottom-up and with a lag for women writh respect to men.

ROBUSTNESS CHECKS. I perform a number of checks to assess the robustness of the long-run analysis. The results are summarized in Figure A16. The baseline regression controls for population (1860), men and female literacy rates (1860), wealth per capita (1855) and local taxes per capita (1855). The baseline point estimate is represented by the horizontal line. I then control for population growth between 1860 and 1900, male and female literacy rates growth between 1860 and 1900, income shocks between 1900 and 1930 (rainfall variability), land inequality (1860), and educational measures (enrollment and school supply in 1908). The estimated coefficient remains very robust to all these additional controls. Next, I control for covariates capturing spatial variation such as latitude and longitude as well as other variables related to the size of municipalities and population dispersion. Although the point estimate falls slightly when controlling for latitude and municipality's size, it remains positive and significant, and it is not significantly different from the

baseline estimate. Last, I perform a number of sample exclusions to ensure that my results are not driven by outliers or only one province. I exclude municipalities with a walking distance to the closest port below 20km, with extreme levels of population (1900), literacy rates (1900) or missing men (1900-1930), and exclude provinces one by one. When excluding the province of Coruña the point estimate increases slightly because the effect is smaller in this province and it accounts for almost one third of all municipalities. Overall, the results remain fairly robust to all these sample restrictions.

OTHER OUTCOMES. I also explore the potential long-run effects of mass emigration on other educational outcomes. In particular, I look at the number of public and private schools, public libraries, and local associations per thousand inhabitants in 2018. As displayed in Table A11, the average share of missing men (1900-1930) did not seem to have increased the supply of public or private schools. These results suggest that despite large investments in schools made by migrants, in the long run the State may have tried to compensate other areas. It is impossible to determine, however, the extent to which migrants might have crowded-out the State. In terms of the availability of public libraries, I find no significant difference between areas that experience more or less migration despite anecdotal evidence that migrants financed the opening of several libraries. Using the registry of all Galician local associations (approximately 20,000), I find that the average share of missing men (1900-1930) reduced the number of local associations per inhabitant, probably due to its effects on population density. However, when looking at the share of associations having an educational goal, I find a large positive impact. This higher prevalence of educational local associations might be the legacy of the hundreds of migrants' associations with that aim that were created one century ago. I discuss this aspect further in the next section.

6 Mechanisms

6.1 Selection into emigration, return migration and incentives

In the previous section, I show that the emigration has a short-run negative impact on literacy of both males and females but just temporary. One decade later emigration leads to net gains in male literacy rates while female literacy rates recovered their baseline level. In this section, I discuss some mechanisms that can explain these results.

Age composition of migrants. The short-run negative effect can be partially explained by the departure of young individuals and the fact that literacy rates in the population decreased with age (especially among women). According to the statistics compiled by Vázquez-González (1999), around 54% of migrants were younger than 25 years old (Table A1). Moreover, at any given point in time, literacy rates were higher among younger individuals (Table A2). For instance, in 1900 close to 20% of women in their 20s were literate compared to 13% in their 40s. By 1930, literacy rates were 62% and 41% among women in their 20s and 40s respectively. As shown in Table A2, this age differential was pronounced among women but smaller among men.

Historical evidence on selection. The degree of selection in literacy among migrants could amplify or soften this age composition effect, however, the scarce evidence available points to no selection conditional on age. The (aggregate) migration statistics suggest that Galician (as well as Spanish) migrants were positively selected but the pattern changes once age is taken into account. For instance, the detailed information compiled by Vázquez-González (1999) shows that 87% of male migrants in the 1920s were literate, compared to 1930 men's literacy rates of 61% on average and 81% among young men (i.e., in their 20s). Table A1 gathers the only available information about Galician migrants' literacy coming from the Official Migration Statistics and the data of Vázquez-González (1999). Official Migration Statistics may over-estimate the degree of literacy since embarkation lists often classified individuals knowing how to sign as literate. An additional reason is that literacy rates would probably be lower among clandestine migrants. Despite the caveats of official statistics, they seem to provide strong evidence that Galician migrants were not negatively selected. We lack information about literacy among returnees, yet the data for Spain shows that migrants and returnees had similar literacy rates suggesting no selection into return migration for the country as a whole.

Descriptive evidence with data by cohort. To explore more in depth the role played by selection into migration, return migration and investments into children's education, I exploit data by cohort from the population censuses. This information is only available at the province level (in every census year) and for the 23 largest municipalities in 1900 and 1910.[31]

The census data for Galicia shows that the size of young cohorts experienced a large decrease

---

[31]Despite this caveat, the distribution of the share of missing men in this sub-sample is very similar to the full sample.

during the mass emigration era. As shown in Figure A18, in 1900 there were around 100 thousand men aged 11 to 15 in Galicia, 20 years later the size of this cohort had decreased to 41 thousand (i.e., a 55% decline). In the case of women, the size of the cohort goes from 100 to 67 thousand (i.e., a 33% decline). Comparing Galicia with the rest of Spain, we observe that between 1920 and 1940 the size of the cohort barely falls in Galicia (7% drop) compared to the rest of the country (20% drop).[32] This points to the compensating effect of return migrants. Note that according to the official statistics as many as 60% of Galician migrants eventually return, a figure similar to that from other Southern European countries (Bandiera et al., 2013). The roe of return migration is a crucial difference from other settings of forced migration in which individuals are not allowed to come back to their communities of origin (e.g., Testa, 2020).

The same pattern emerges using the data across districts between 1900 and 1910. There is a strong negative relationship between the reduction in size of the cohort born between 1885 and 1889 and the increase in the share of missing men in 1900-1910 (Figure A19). Similarly, changes in average literacy rates in this cohort are also correlated with changes in missing men, suggesting that young individuals were investing more in education in high migration areas (Figure A20). Although this descriptive evidence is only suggestive, it goes in line with the historical evidence highlighting that many migrants invested in education before departing (Vázquez-González, 1999; Gabriel, 2013) and with other with other studies showing that remittances and the option to emigrate can increase parental investments on children's educatino (Dinkelman & Mariotti, 2016). Moreover, the fact that there is an increase in both boys and girls literacy suggests that reason behind the null medium-run effect on women's education is not a lack of investments on them (Shresta, 2008), but rather related to the role of return migration and selection into return.

6.2 Migrants' associations and schools, 1900-1940

Background. In clear contrast with episodes of forced migration (Becker and Ferrara, 2019), Galician migrants maintained strong links with their communities of origin and founded hundreds of associations with the aim of financing investments in their hometowns. Most of these associations had an educational goal, such as purchasing materials or funding the construction of new schools in their hometowns. Migrants believed that these institutions were the best way in which they could contribute to the development of Galicia in the long run (q. 7, 8, 10, 19, 20, 21). The schools constructed were of outstanding quality for the standards of the time: large, bright, fully

---

[32]The evolution of the cohort aged 6 to 10 in 1900 displays a very similar pattern during these decades.

equipped, etc. Migrants even stressed the importance of hygiene and vaccination (q. 16, 24) and offered free resources to poor students (q. 28, 29).[33] Moreover, these schools provided high quality education manifested by novel teaching practices (q. 30, 31) and a different curriculum (e.g., no religious education, courses on geography, accounting, farming). Some critics argued that migrants' schools fostered even more migration and that children were not educated into the Galician tradition (q. 32, 33, 34, 35). The predominant view however seems to be that of praising these investments and aspiring to have migrants' schools in every single municipality (q. 7, 22, 23). As shown in Figure A22, most associations were founded between 1910 and and 1924 while most schools opened between 1920 and 1934. Figure A23 and Figure A24 show the location of these schools over time and the strong overlap between these investments and the extent of missing.

Foundation of migrants' associations and schools. To better understand the determinants behind the foundation of migrants' associations and schools, I perform cross-sectional regressions with key historical variables. As in the long-run analysis, I instrument the share of missing men (1900-1930) with the adult sex ratio in 1860. Table 6 summarizes the results. Column (1) shows that most variables are not significant predictors of the existence of at least one association. Interestingly, the share of missing is not a relevant factor at this margin. Both higher taxes per capita (1855) and the availability of more schools (1875) reduce the likelihood that one association was created in the municipality. One of the reasons why it may be so difficult to predict the opening of an association is because some of them originated in a municipality but operated in several, and hence, have been assigned to all those locations. In Column (2), the dependent variable is the number of associations. The share of missing men (1900-1930) seems to be a strong determinant of migrants' associations, although the coefficient is not significant (p-value=0.13). An increase in the share of missing men (1900-1930) in on standard deviation translates into one more migrant associations on average. Larger municipalities and those closer to a transatlantic port also had significantly more migrants' associations while municipalities with higher taxes per capita (1855) and more schools (1875) significantly fewer. An increase in children's enrollment rates (1875) in one standard deviation is associated with an increase of 0.3 units in the number of associations. In columns (3) and (4) the dependent variable is the number of schools founded by migrants. The last column includes the number of migrants' associations as an additional variable. Focusing on column (4), the major determinants of the number of migrants' schools are the share of missing

---

[33]Figure A21 shows some of them.

men (1900-1930), the number of migrants' associations, local shocks (i.e., rainfall variability) and literacy rates. Even after controlling for the number of associations, an increase in the share of missing men (1900-1930) in one standard deviation leads to 1.6 more schools on average. This results suggest that the size of the migrant community abroad was crucial for the financing of new schools. Importantly, economic forces also played a role. Land inequality (measured by the share of landless peasants) and local economic shocks (rainfall variability) are negatively correlated with the number of new schools. For instance, one standard deviation increase in rainfall variability is associated with 0.75 less schools on average. The large and negative coefficient of rainfall variability may indicate that migrants faced a trade-off between individual and collective remittances. If families were in need of resources to face these shocks, it would have been more difficult that migrants could collect money to finance new interventions.

Impact of migrants' associations and schools. Table 7 presents the results of the cohort analysis. In columns (1) to (3) the dependent variable is the share of individuals with completed primary education while in columns (4) to (6) the ratio between the size of cohort c and $c - 1$. Column (1) shows that the opening of an association does not seem to have any impact on children's primary achievement. Yet, this effect may be confounded by the fact that most associations financed a new school 5 to 10 years later. A new migrant school, in fact, seems to increase primary completion among the children exposed (column 2). When both variables are considered together, as in column (3), we see that foundation of associations does have a negative effect on educational attainment while the opening of a school a positive one. The point estimate of column (3) implies that a new school increases the primary completion rate by 1 percentage point among children younger than 10. This effect is modest (i.e., 4% of a standard deviation) but one needs to consider that often several schools opened in the same municipality.

Looking at the size of the cohorts (columns 4-6), a similar pattern emerges. The foundation of a migrant association reduces the size of young cohorts while the opening of a school has the reverse effect. As shown in column (6), the creation of an association is associated with a fall of 3.3 percentage points in the size of the cohort of children younger than 10, although the effect is imprecisely estimated. A new migrant school, however, increases the size of the cohort by 1.7 percentage points (i.e., 3% of a standard deviation). Note that this effect may be somewhat mechanical if the foundation of an association is preceded by a large migration flow that in turn reduces fertility. However, there are significant differences between men and women what goes

against this hypothesis.

The heterogeneous effects by gender may shed some light on these results. I find that the positive impact of a new migrant school on educational attainment is very similar for both men and women. The negative effect of the foundation of an association however is considerably larger for women (the point estimate is twice as large although not statistically different). This result is consistent with the hypothesis that migrants' association may have encouraged family migration (and/or reunification) therefore affecting more women than men.

6.3 Beliefs about education and effort

Background. The work of historians suggests that Galician migrants realized the value of education once at Latin America and tried to transmit these norms (Peña Saavedra, 1983; De Gabriel, 2012). There is plenty of anecdotal evidence in this direction, some of which is shown in Section G. Migrants acknowledged their lack of education and the importance of it (q. 1, 2, 3), not just in the face of migration (q. 3, 4, 6, 8) but for the development of Galicia (q. 2, 3, 5, 7, 8). The diffusion of these values for education would take place through migrants' associations and their investments (q. 7, 8, 10, 19, 20 , 21, 22, 23, 24), migrants' missives, magazines and other publications (q. 9, 13, 14) as well as with the return of migrants and the transmission from parents to children (q. 11, 12, 13). Migrants not only encouraged directing funds towards educational resources (q. 18), but stressed the importance of regularly attending school (q. 12, 14, 15) and urged public authorities to enforce compulsory education laws (q. 16, 17). Moreover, migrants worked extremely hard at destination, saving most of their earnings, what could have also contributed to change their mindset about the value of effort (q. 26, 25, 27).

Empirical analysis. To test whether emigration could have led to a change in values about the importance of education, I combine my historical data with contemporary survey data based on the municipality of respondents. I use several waves of the Encuestas Opinión Pública where individuals report in a scale from 0 to 10 whether success in life is due to one's education/effort (rather than luck/contacts). I estimate the following equation by 2SLS:

$$V_{ij} = \alpha + \gamma \overline{ShMissing_i} + X_i \Pi + Z_j \Psi + u_{ij} \quad (8)$$

where V_{ij} is the value individual i (in municipality j) puts on education and effort, $\overline{ShMissing_i}$ is the average share of missing men (1900-1930) in municipality j, X_i and Z_j stand for individual

and municipality level controls controls and u_{ij} represents standard errors that I cluster at the municipality level. I instrument $\overline{ShMissing_i}$ with the proxy for pioneer migrant networks (i.e., adult sex ratio in 1860).

The results, displayed in Table 8, show that individuals living in municipalities that experienced higher historical emigration put greater value on education and effort. On average, one standard deviation increase in the share of missing men (1900-1930) is associated with a 22% standard deviation increase in the reported valuation (e.g., around 0.5 points in a 0-10 scale with an average of 5). The results are very robust to the inclusion of individual and historical controls.[34] I do not find any impact on other related dimensions such as their satisfaction with public education, their willingness to pay more taxes in exchange of more public goods, nor their sense of how much society benefits from taxes (Table A12). I do find, however, a small reduction in the share of people declaring that we spend too little on education or that we pay too little taxes.

7 Conclusion

This paper examines the impact of mass emigration on human capital accumulation at origin over a century. I show that the Galician mass emigration reduced literacy rates in the short run but its impact became positive after one decade and led to gains in human capital that persist even one hundred years later. I provide empirical evidence that migrants' associations and the schools they financed in their hometowns fostered human capital formation and that migration contributed to shaping social norms at origin about the value of education and effort.

My findings have important implications for both research on migration and public policy. First, they highlight the relevance of taking a long-term view, as the impact of migration and the individuals affected may change over time. For instance, while in the first few decades emigration only affect men's educational attainment, in the long run both men and women benefited similarly. They also stress the need to consider other mechanisms beyond monetary remittances, such as migrants' financing of public goods in their hometowns and social remittances. My results suggest that fostering migrants engagement with their communities of origin (e.g., through local associations, information campaigns, fundraising) can be an effective way to promote human capital accumulation and economic development at origin. Innovative public policies could be implemented to strengthen these links and encourage the transmission of collective remittances

---

[34]One would expect that controlling for individuals' education the effect of the share of missing men (1900-1930) should decrease. One reason why it may not be the case is that educational attainment is measured in categories and potentially with error (e.g., attended vs. completed studies).

targeting specific goals. Finally, my findings shed new light on the impact of migration on the adoption of new social norms, a channel that may lead to persistent effects in the long run. Future work should examine in more depth the interplay between these different mechanisms, and explore the long-run impact of emigration on other social and economic aspects in the communities of origin.

References

ABRAMITZKY, R., BOUSTAN, L. P., AND ERIKSSON, K. 2012. Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration. *American Economic Review* 102:1832–56.

ANDERSSON, D., MOUNIR, K., AND PRAWITZ, E. 2017. Mass migration, cheap labor, and innovation. *working paper* .

ANDREWS, I., STOCK, J. H., AND SUN, L. 2019. Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics* 11:727–753.

ANTMAN, F. M. 2012. Gender, educational attainment, and the impact of parental migration on children left behind. *Journal of Population Economics* 25:1187–1214.

BADA, X. 2014. Mexican hometown associations in Chicagoacán: From local to transnational civic engagement. Rutgers University Press.

BATISTA, C., LACUESTA, A., AND VICENTE, P. C. 2012. Testing the 'brain gain'hypothesis: Micro evidence from cape verde. *Journal of Development Economics* 97:32–45.

BAZZI, S., FISZBEIN, M., AND GEBRESILASSE, M. forthcoming. Frontier culture: The roots and persistence of" rugged individualism" in the united states. Technical report.

BECKER, S. O. AND FERRARA, A. 2019. Consequences of forced migration: A survey of recent findings. *Labour Economics* 59:1–16.

BECKER, S. O., GROSFELD, I., GROSJEAN, P., VOIGTLANDER, N., AND ZHURAVSKAYA, E. 2020. Forced migration and human capital: evidence from post-wwii population transfers. *American Economic Review* 110:1430–63.

BERTRAND, J. R. 1999. La emigración gallega,¿ sistema original? *SEMATA, Ciencias Sociais e Humanidades* 11:439–455.

BOLT, J., INKLAAR, R., DE JONG, H., AND VAN ZANDEN, J. Maddison project database, version 2018. *Rebasing 'Maddison': new income comparisons and the shape of long-run economic development* .

BRUM, M. 2019. Italian migration to the united states: The role of pioneers' locations. *working paper* .

CALVI, R., MANTOVANELLI, F., AND HOEHN-VELASCO, L. forthcoming. The protestant legacy: Missions and human capital in india. *Journal of Human Resources* .

CAMPOS ALVAREZ, J. R. 1993. La emigración gallega a américa, 1880-1930. integración y retorno. *Minius* pp. 133–145.

CHAUVET, L., GUBERT, F., MERCIER, M., AND MESPLÉ-SOMPS, S. 2015. Migrants' home town associations and local development in mali. *The Scandinavian Journal of Economics* 117:686–722.

CHRISTIAN, P. AND BARRETT, C. B. 2017. Revisiting the effect of food aid on conflict: A methodological caution. The World Bank.

CLEMENS, M. A. AND MENDOLA, M. 2020. Migration from developing countries: Selection, income elasticity, and simpson's paradox. *IZA Discussion Paper* .

CONNOR, D. S. 2019. The cream of the crop? geography, networks, and irish migrant selection in the age of mass migration. *The Journal of Economic History* 79:139–175.

DÍEZ-MINGUELA, A., MARTINEZ-GALARRAGA, J., AND TIRADO-FABREGAT, D. A. 2016. Why did spanish regions not converge before the civil war? agglomeration economies and (regional) growth revisited. *Revista de Historia Economica-Journal of Iberian and Latin American Economic History* 34:417–448.

DINKELMAN, T. AND MARIOTTI, M. 2016. The long-run effects of labor migration on human capital formation in communities of origin. *American Economic Journal: Applied Economics* 8:1–35.

DOCQUIER, F. AND RAPOPORT, H. 2012. Globalization, brain drain, and development. *Journal of Economic Literature* 50:681–730.

DROLLER, F. 2018. Migration, population composition and long run economic development: Evidence from settlements in the pampas. *The Economic Journal* 128:2321–2352.

DUFLO, E. 2001. Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American economic review* 91:795–813.

DUSTMANN, C. AND WEISS, Y. 2007. Return migration: theory and empirical evidence from the uk. *British Journal of Industrial Relations* 45:236–256.

EIRAS ROEL, A. 1991. La emigración gallega a américa: panorama general. *In* La emigración española a Ultramar, 1492-1914, pp. 17–40. Asociación de Historia Moderna.

EYNDE, O. V. 2016. Military service and human capital accumulation: Evidence from colonial punjab. *Journal of Human Resources* 51:1003–1035.

FARÍAS, R. 2010. La inmigración gallega en el sur del gran buenos aires:(1869-1960).

FERNÁNDEZ-SANTIAGO, M. X. 2008. Asociacionismo gallego en américa, 1871-1960. *ESPAÑOLA A AMÉRICA* p. 199.

FOX, J. AND BADA, X. 2008. Migrant organization and hometown impacts in rural mexico. *Journal of Agrarian Change* 8:435–461.

GABRIEL, N. D. 2006. Ler e escribir en galicia. a alfabetización dos galegos e das galegas nos séculos xix e xx.

GABRIEL, N. D. 2013. El proceso de alfabetización en galicia: un intento de explicación y comprensión. *Historia de la Educación* 32:289–313.

GOLDSMITH-PINKHAM, P., SORKIN, I., AND SWIFT, H. 2020. Bartik instruments: What, when, why, and how. Technical report.

GROSJEAN, P. AND KHATTAR, R. 2018. It's raining men! hallelujah? the long-run consequences of male-biased sex ratios. *The Review of Economic Studies* .

HATTON, T. J. AND WILLIAMSON, J. G. 1998. The age of mass migration: Causes and economic impact. Oxford University Press on Demand.

HERNÁNDEZ BORGE, J. 1986. Notas sobre la emigración gallega en la segunda mitad del siglo xix.

JAEGER, D. A., RUIST, J., AND STUHLER, J. 2018. Shift-share instruments and the impact of immigration. Technical report, National Bureau of Economic Research.

KARADJA, M. AND PRAWITZ, E. 2019. Exit, voice, and political change: Evidence from swedish mass migration to the united states. *Journal of Political Economy* 127:000–000.

KELLY, M. 2019. The standard errors of persistence. *CEPR Discussion Paper No. DP13783* .

KNUDSEN, A. S. B. 2020. Those who stayed: Individualism, self-selection and cultural change during the age of mass migration. *working paper* .

MCKENZIE, D. AND RAPOPORT, H. 2010. Self-selection patterns in mexico-us migration: the role of migration networks. *The Review of Economics and Statistics* 92:811–821.

MCKENZIE, D. AND RAPOPORT, H. 2011. Can migration reduce educational attainment? evidence from mexico. *Journal of Population Economics* 24:1331–1358.

NÚÑEZ-SEIXAS, X. M. 1994. Migraciones y política. las remesas invisibles. algunas notas sobre la influencia socio-política de la emigración transoceánica en galicia,(1890-1930). *Estudios Migratorios Latinoamericanos* 9:301–346.

NÚÑEZ-SEIXAS, X. M. 1998a. Emigrantes, caciques e indianos: o influxo sociopolítico da emigración transoceánica en Galicia (1900-1930). Xerais de Galicia.

NÚÑEZ-SEIXAS, X. M. 1998b. Retornados e inadaptados: el" americano" gallego, entre mito y realidad (1880-1930). *Revista de Indias* 58:555–593.

NÚÑEZ-SEIXAS, X. M. 2001. La Galicia austral: la inmigración gallega en la Argentina, volume 3. Editorial Biblos.

NÚÑEZ-SEIXAS, X. M. AND VÁZQUEZ-SOUTELO, R. 2005. As cartas do destino: unha familia galega entre dous mundos, 1919-1971. Editorial Galaxia.

PALMÁS, R. 1976. La emigración. *FABRA, Gustavo, et al. Los gallegos. Madrid, Istmo* .

PEÑA-SAAVEDRA, V. 1991. Las escuelas de americanos y las sociedades de instrucción en galicia. *PhD thesis, Universidade de Santiago de Compostela, Santiago de Compostela* 1-4.

PEÑA-SAAVEDRA, V. 1995. As sociedades galegas de instrucción: proxecto educativo e realizacións escolares. *Estudios migratorios* pp. 8–83.

PEÑA-SAAVEDRA, V. 2010. Universidade de Vigo.

RAPOPORT, H., SARDOSCHAU, S., AND SILVE, A. 2020. International migration and cultural convergence. Technical report.

ROCHA, R., FERRAZ, C., AND SOARES, R. R. 2017. Human capital persistence and development. *American Economic Journal: Applied Economics* 9:105–36.

ROSÉS, J. R., MARTÍNEZ-GALARRAGA, J., AND TIRADO, D. A. 2010. The upswing of regional income inequality in spain (1860–1930). *Explorations in Economic History* 47:244–257.

SÁNCHEZ-ALONSO, B. 1995. Las causas de la emigración española, 1880-1930. Alianza Editorial.

SÁNCHEZ-ALONSO, B. 2000. Those who left and those who stayed behind: explaining emigration from the regions of spain, 1880–1914. *The Journal of Economic History* 60:730–755.

SEQUEIRA, S., NUNN, N., AND QIAN, N. 2020. Immigrants and the making of america. *Review of Economic Studies* .

SHRESTHA, S. A. 2017. No man left behind: Effects of emigration prospects on educational and labour outcomes of non-migrants. *The Economic Journal* 127:495–521.

SHRESTHA, S. A. AND PALANISWAMY, N. 2017. Sibling rivalry and gender gap: intrahousehold substitution of male and female educational investments from male migration prospects. *Journal of Population Economics* 30:1355–1380.

SPITZER, Y. AND ZIMRAN, A. 2018. Migrant self-selection: Anthropometric evidence from the mass migration of italians to the united states, 1907–1925. *Journal of Development Economics* 134:226–247.

STOLZ, Y. AND BATEN, J. 2012. Brain drain in the age of mass migration: Does relative inequality explain migrant selectivity? *Explorations in Economic History* 49:205–220.

TESTA, P. forthcoming. Migration from developing countries: Selection, income elasticity, and simpson's paradox. *The Economic Journal* .

VALENCIA CAICEDO, F. 2019. The mission: Human capital transmission, economic persistence, and culture in south america. *The Quarterly Journal of Economics* 134:507–556.

VÁZQUEZ-GONZÁLEZ, A. 1999. La emigración gallega a américa, 1830-1930. *Memoria de doctorado inédita, Facultade de Ciencias Económicas e Empresariais, Universidade de Santiago de Compostela, Santiago de Compostela* 1-4.

VIANA, V. 2017. La emigración a las Américas por el Puerto de Vilagarcía de Arousa. Teofilo Edicións.

Figures

FIGURE 1: Rainfall in Galicia, 1800-1930
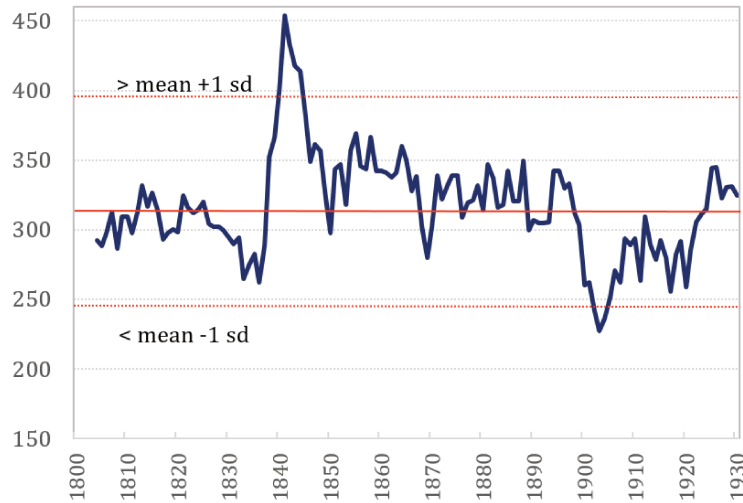
Source: Pauling et al. (2006) European Gridded Seasonal Precipitation Reconstructions.
Note: The figure shows the 5-year moving average of precipitation in Galicia during the Fall season. Return to page .

FIGURE 2: Rainfall shocks and adult sex ratio (1860)
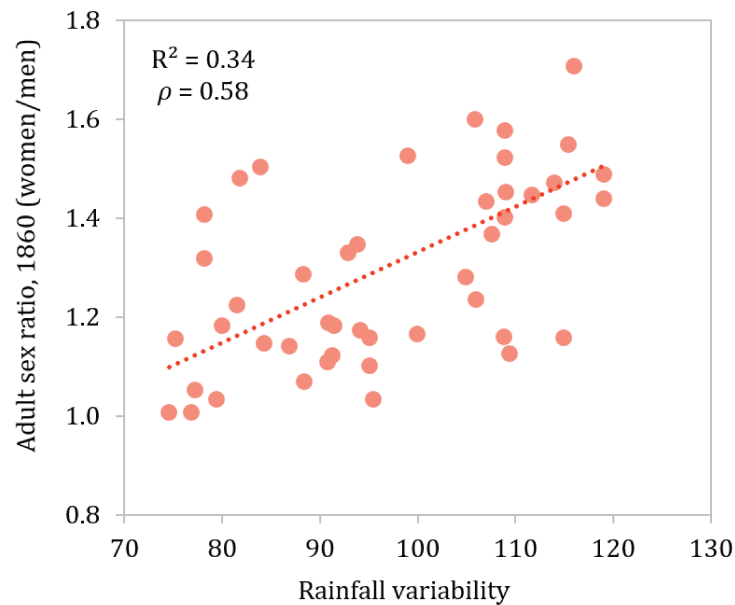
Note: Each dot represents a district. Rainfall information is available at the 0.5°*x* 0.5°(24 cells in Galicia). The measure of rainfall variability represents the standard deviation of precipitation in the Fall season between 1830 and 1860. Return to page .

FIGURE 3: Main migrant destinations across Galician municipalities, 1900-1930

Rio de la Plata

0.00 - 0.50
0.50 - 0.70
0.70 - 0.85
0.85 - 1.00

Cuba

0.00 - 0.07
0.07 - 0.20
0.20 - 0.40
0.40 - 1.00

Brazil

0.00 - 0.02
0.02 - 0.10
0.10 - 0.20
0.20 - 1.00
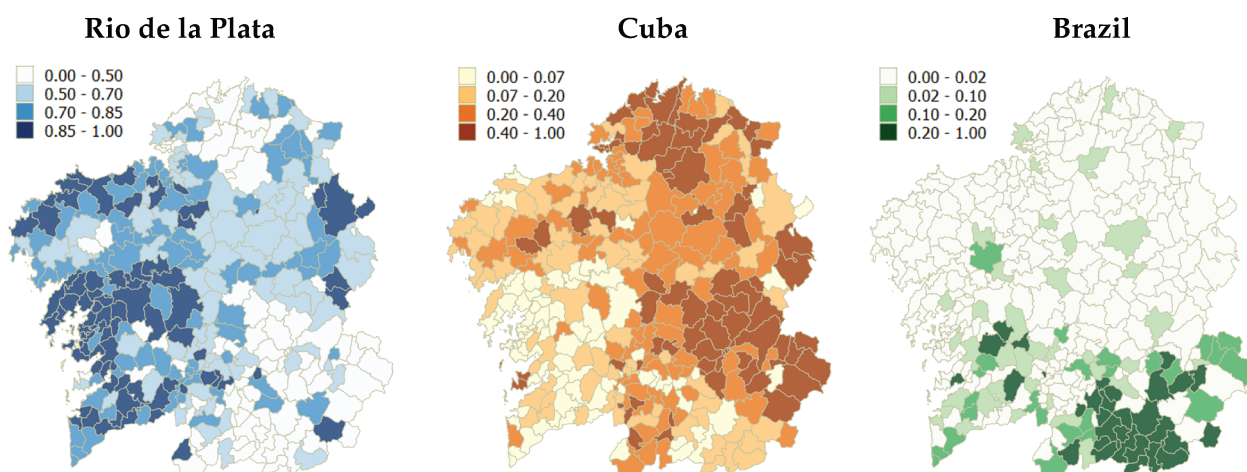
Source: Adapted from the records compiled by Vázquez-González (1999).
Note: The map shows the share of migrants going to each of the three main destinations between 1900-1930. Rio de la Plata stands for Argentine and Uruguay. Return to page 8 or 14.

FIGURE 4: GDP per capita growth in the main destinations, 1877-1930
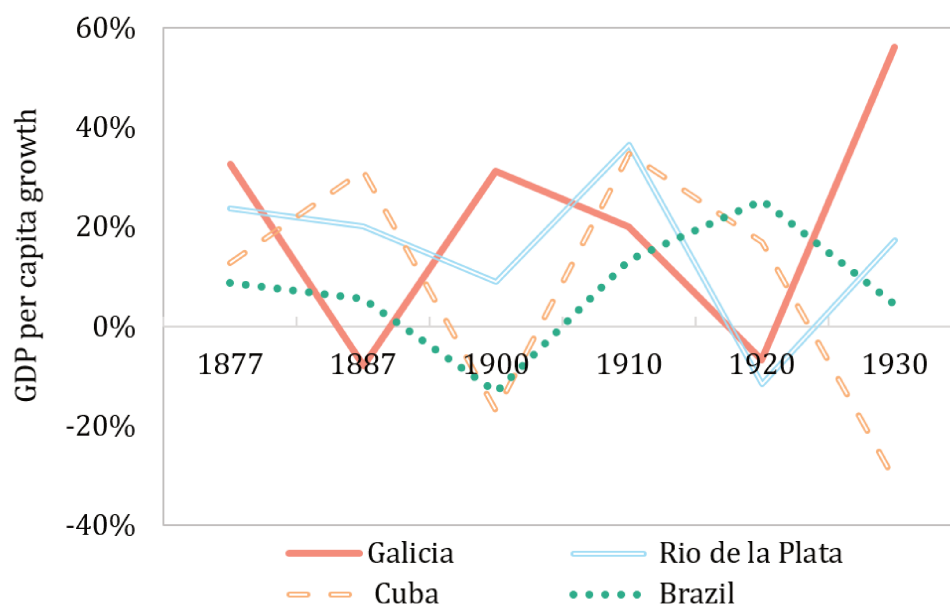
Source: Maddison Project Database 2018 for Latin American countries and Spain, Rosés et al. (2010) for Galicia (relative to Spain). *Note*: The figure displays GDP per capita growth by decade. Return to page 8 or 14 .

FIGURE 5: Emigration and missing men in Galicia, 1877-1930
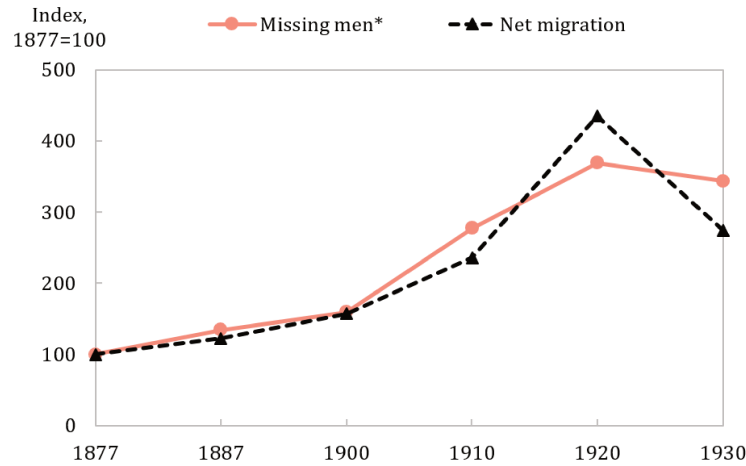
Source: Vázquez-González (1999) and Spanish population census (1877-1930).
Note: "Net migration" stands for sea departures minus arrivals in the last 15 years. "Missing" stands for residents absent the day of the census. Transients were subtracted from the missing to avoid double counting. Return to page 10.

FIGURE 6: Emigration and missing men across Galician provinces, 1900-1920
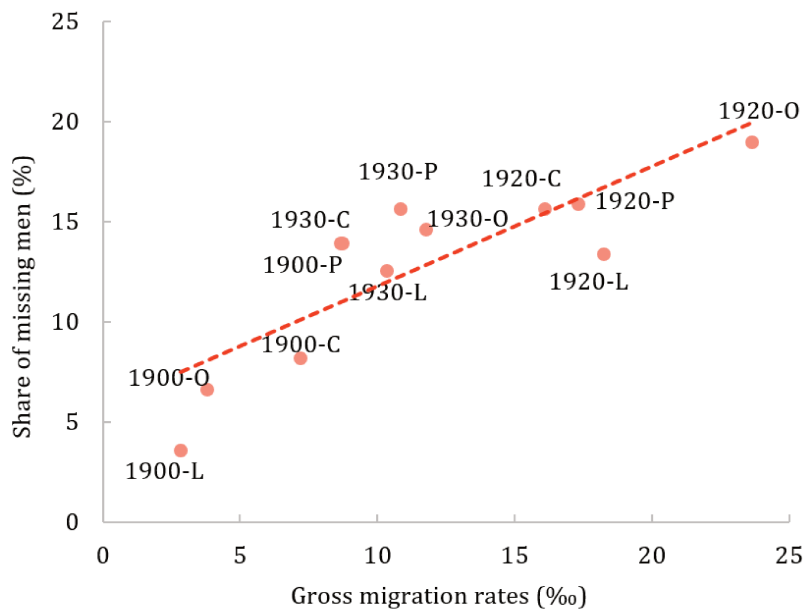
Source: Vázquez-González (1999) and Spanish population census (1900-1920).
Note: "Migration" stands for sea departures. "Missing" stands for residents absent the day of the census. Transients were subtracted from the missing to avoid double counting. C: Coruña, L: Lugo, O: Ourense, P: Pontevedra. Return to page 10.

FIGURE 7: Adult sex ratio in 1860
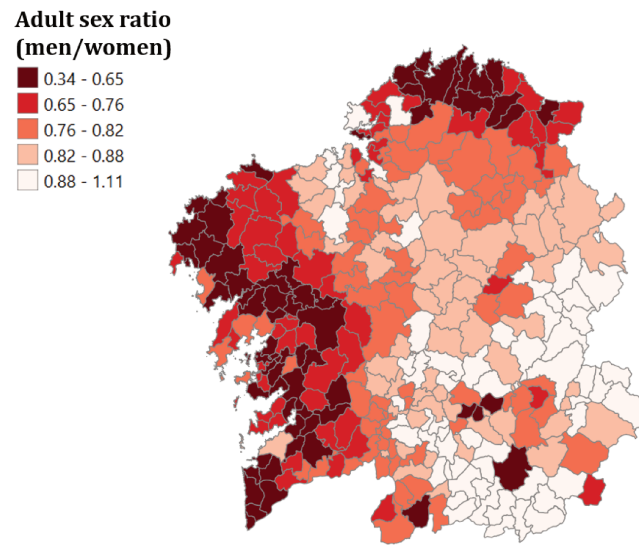
Note: The figure shows the adult sex ratio in 1860 computed as the ratio between men and women aged 30 to 60. Return to page 11.

FIGURE 8: Share of missing men across Galician municipalities (%), 1877-1930
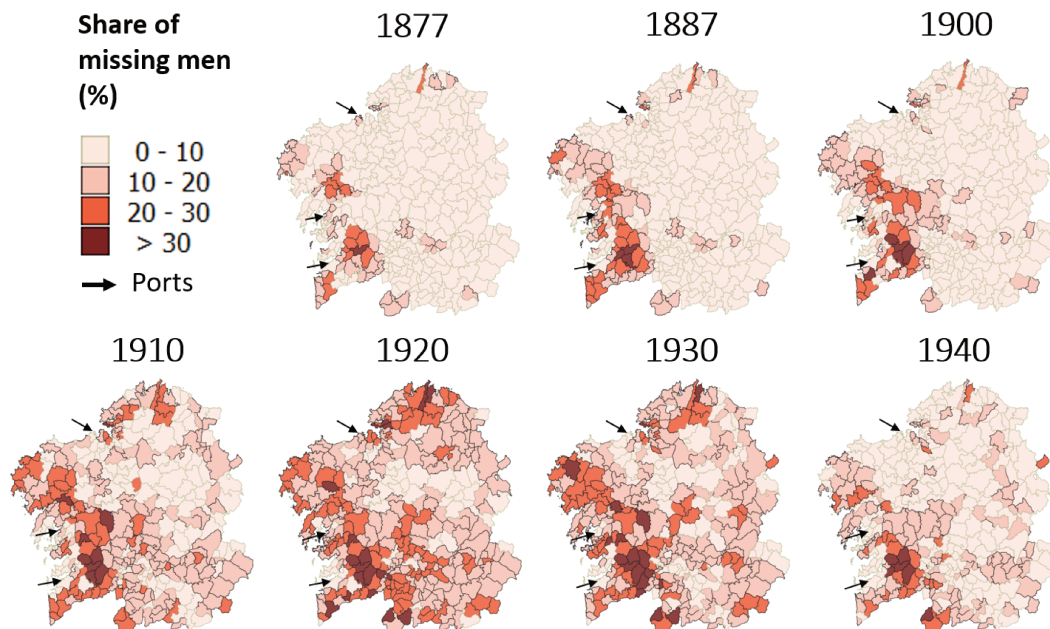
Source: Spanish population census (1877- 1930).
Note: The figure shows the share of missing men the day of the census in different years. The arrows indicate the ports shipping boats to Latin America. Return to page 17.

FIGURE 9: Evolution of literacy rates in Galicia, 1860-1940
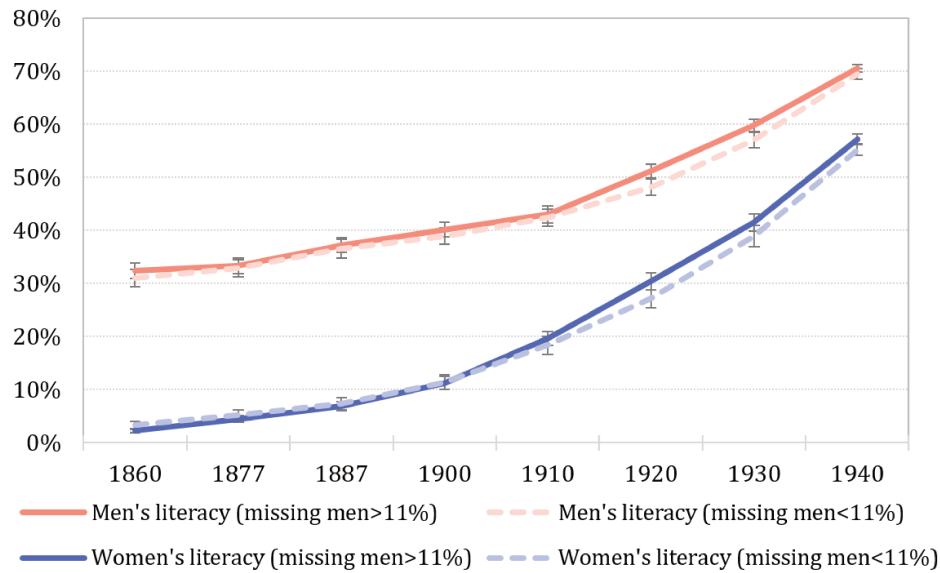
Source: Spanish population census (1860- 1940). *Note*: The solid (/dashed) lines represent the average for municipalities with an average share of missing men (1900-1930) above (/below) the median. Return to page 17.

FIGURE 10: Gap in literacy rates and share of missing men, 1860-1940
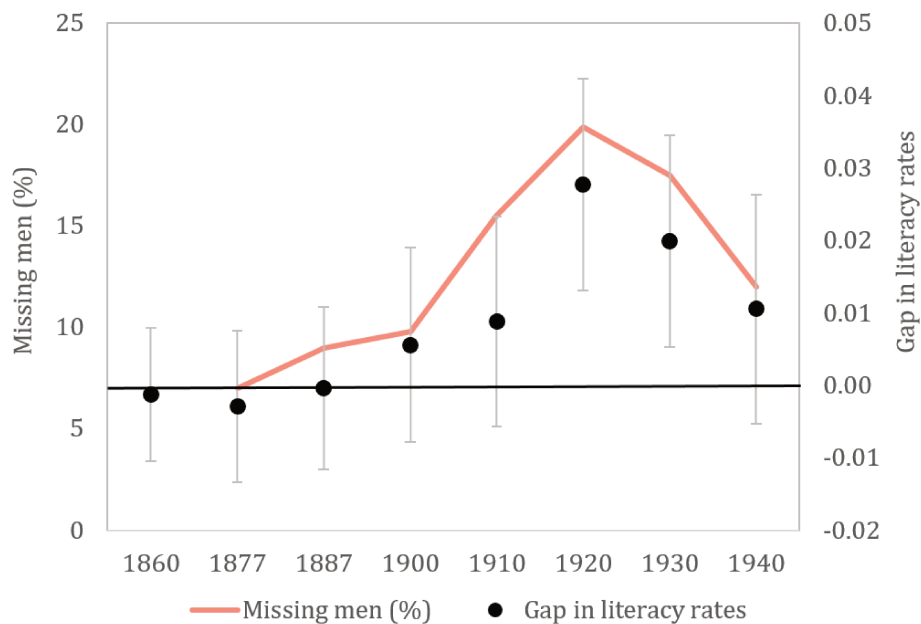
Source: Spanish population census (1860- 1940).
Note: The figure shows the share of missing men in Galicia (left) and the gap in literacy rates between municipalities with an average share of missing men (1900-1930) above and below the median (right). Error bars represent 95% confidence intervals of the difference in means. Return to page 17.

FIGURE 11: Average impact of missing men on literacy rates, 1877-1930
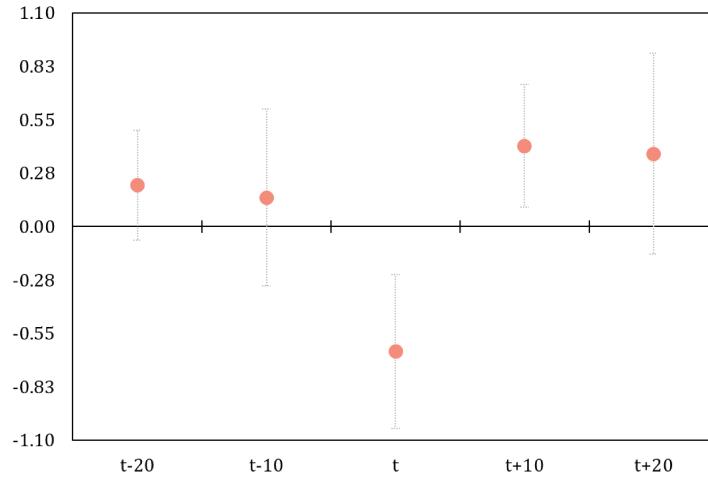
Note: The figure plots β coefficients from estimating variants of equation (1) with different lags of the share of missing men or literacy rates. For instance, the estimate of *t-k* is obtained from estimating the impact of the share of missing men in year *t* on literacy rates in year *t*-k in a panel of 4 periods (decades). The estimate of $t + k$ is obtained from estimating the impact of missing men in *t-k* on literacy rates in *t* in a panel of 4 periods (decades). The share of missing men (in year *t*) is instrumented by the growth in GDPpc at the average migrant destination (between *t*-10 and *t*) and its interaction with the adult sex ratio in 1860. Dotted lines indicate 95% confidence intervals around the point estimates. Return to page 20.

FIGURE 12: Impact of missing men on literacy rates, 1900-1930
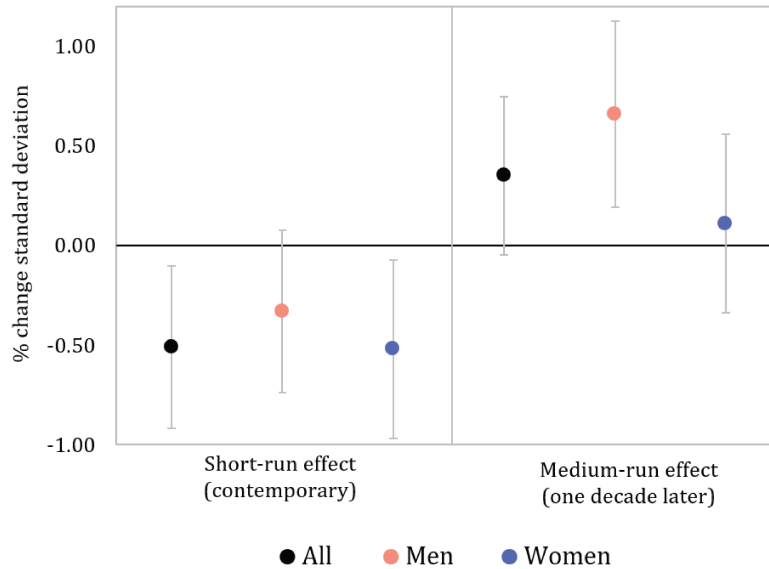
Note: The figure shows β coefficients from estimating equations (1) and (2) by 2SLS. The short-run effect suggests that one standard deviation increase in the share of missing men (in year *t*) led to a drop in literacy rates in about 51% of a standard deviation The medium-run effect suggests that one standard deviation increase in the share of missing men (in year *t*) led to an increase in literacy rates (10 years later) in about 35% of a standard deviation The share of missing men (in year *t*) is instrumented by the growth in GDPpc at the average migrant destination and its interaction with the adult sex ratio in 1860. Dotted lines indicate 95% confidence intervals around the point estimates. Return to page 20.

FIGURE 13: Long-run impact of missing men on the share of secondary education
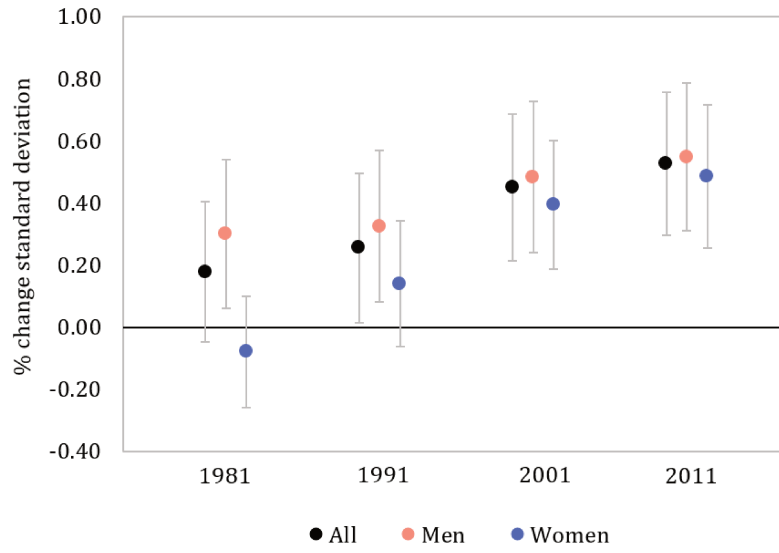
Note: The figure plots β coefficients from estimating equation (4) by 2SLS in different years. The estimates of the long-run effect suggest that one standard deviation increase in the average share of missing men (1900-1930) led to an increase in the share of individuals with completed secondary education in about 50% of a standard deviation in 2011. The average share of missing men (1900-1930) is instrumented by the adult sex ratio in 1860. Return to page 22.

Tables

TABLE 1: Departures from Europe to the Americas, 1850-1930

	Departures (m.),	Gross migration rates (departures/1000 people)						
	1815-1930	1861-1870	1871-1880	1881-1890	1891-1900	1901-1910	1911-1920	1921-1930
England	11.4	2.80	4.00	5.60	3.60	5.50	7.60	2.70
Italy	9.9	1.10	3.50	5.00	10.80	16.30	3.40	
Ireland	7.3	14.60	6.60	14.20	8.90	7.00	6.80	5.90
Austria-Hungary	5.0	–	–	1.10	1.60	4.80	6.10	1.40
Germany	4.8	–	1.50	2.90	1.00	0.50	0.40	1.00
Spain	4.4	–	–	3.40	3.40	7.00	10.60	4.00
Portugal	1.8	1.90	2.90	4.30	5.60	6.50	13.90	5.30
Galicia	**1.5**	**1.40**	**4.16**	**5.82**	**7.14**	**14.62**	**25.38**	**14.52**
Sweden	1.2	3.10	2.40	7.00	4.10	4.20	3.10	1.80
Norway	0.8	5.80	4.70	9.50	4.50	8.30	4.20	3.10
Finland	0.4	–	–	1.30	2.30	5.50	6.40	2.10
France	0.4	0.20	0.20	0.30	0.10	0.10	0.20	–
Denmark	0.4	–	2.10	3.90	2.20	2.80	3.20	1.70
Switzerland	0.3	–	1.30	3.20	1.40	1.4	1.70	1.40
Netherlands	0.2	0.60	0.50	1.20	0.50	0.50	0.40	0.50
Belgium	0.2	–	–	0.90	0.40	0.60	1.00	0.30

Source: Baines (1991), Sánchez-Alonso (1995) and Vázquez-González (1999).

Note: Gross migration rates represent yearly departures per thousand people at baseline. For instance, a rate of 10 means that every year there were departures equivalent to 1% of the baseline population. Return to page 7.

TABLE 2: Explaining the evolution of missing men, 1887-1930 [Panel first stage]

	Share of missing men (t)				
	(1)	(2)	(3)	(4)	(5)
Average growth at destination	0.07***	0.07**	0.09***	0.07**	0.06*
	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
Average growth at destination x Adult sex ratio (1860)	0.04***	0.04***	0.03**	0.04***	0.04***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.02]
Population in the province (10 years before)		-0.27***			-0.51***
		[0.10]			[0.14]
Literacy rate in the province (10 years before)		0.37**			0.58***
		[0.18]			[0.22]
Missing men (10 years before)			0.24***		0.24***
			[0.03]		[0.04]
Rainfall variability (previous decade)				-0.04	-0.14***
				[0.04]	[0.05]
Municipality and year FE	Yes	Yes	Yes	Yes	Yes
KP F -statistic	9.29	9.48	9.95	8.97	7.80
Mean dependent variable	0.13	0.13	0.13	0.13	0.13
Standard dependent variable	0.09	0.09	0.09	0.09	0.09
Observations (municipalities=322, years=5)	1610	1610	1610	1610	1610

Note: The table shows β coefficients from regressing the average share of missing men (in year *t*) on different lagged controls (equation 3). As an illustration, column (1) shows that one standard deviation increase in the average growth at destination is associated with a 7% standard deviation raise in the share of missing men on average. Increasing the adult sex ratio (1860) by one standard deviation magnifies this effect by 4% more (i.e., a 11% of standard deviation increase in the share of missing men). Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels.

TABLE 3: Explaining the intensity of missing men (1900-1930) [Cross-section first stage]

	Average share of missing men, 1900-1930				
	(1)	(2)	(3)	(4)	(5)
Adult sex ratio (w/m ,1860)	0.34***	0.31***	0.32***	0.35***	0.28***
	[0.07]	[0.07]	[0.06]	[0.06]	[0.06]
Altitude		0.21**			0.12
		[0.10]			[0.11]
Municipality size		-0.24***			-0.20**
		[0.09]			[0.08]
Population settlements		0.04			0.03
		[0.10]			[0.09]
Access to sea		-0.37*			-0.29
		[0.20]			[0.18]
Access to train		-0.09			-0.19
		[0.16]			[0.16]
Distance to port		-0.31**			-0.29**
		[0.13]			[0.13]
Distance to capital		0.09			0.09
		[0.09]			[0.08]
Population (1860)			-0.02		0.01
			[0.08]		[0.07]
Men's literacy (1860)			0.20***		0.17***
			[0.06]		[0.06]
Women's literacy (1860)			-0.11*		-0.11**
			[0.06]		[0.05]
Wealth per capita (1855)				-0.01	0.02
				[0.10]	[0.09]
Local contributions per capita (1855)				0.20*	0.09
				[0.09]	[0.09]
Land inequality (1860)				0.18**	0.20***
				[0.08]	[0.08]
KP F-statistic	23.64	21.93	26.18	33.58	22.47
Mean dependent variable	0.14	0.14	0.14	0.14	0.14
Standard deviation dependent variable	0.10	0.10	0.10	0.10	0.10
Adjusted R-squared	0.19	0.21	0.21	0.20	0.24
Observations	321	321	321	321	321

Note: The table shows β coefficients from regressing the average share of missing men (1900-30) on different covariates (equation 6). As an illustration, column (1) shows that one standard deviation increase in the adult sex ratio (1860) is associated with a 34% standard deviation raise in the average share of missing men (1900-1930). Clustered standard errors at the district level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels.

TABLE 4: Effect of missing men on literacy rates, 1900-1930

| | Average literacy rates (t) | | | |
| | OLS | | 2SLS | |
	(1)	(2)	(3)	(4)
Share of missing men (t)	0.01		-0.51**	
	[0.03]		[0.21]	
Share of missing men (10 years before)		0.07**		0.35*
		[0.03]		[0.21]
Municipality and year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
KP F -statistic	-	-	9.92	15.80
Mean dependent variable	0.35	0.35	0.35	0.35
Standard deviation dependent variable	0.13	0.13	0.13	0.13
Mean share of missing men	0.14	0.11	0.14	0.11
Standard deviation share of missing men	0.08	0.09	0.08	0.09
Observations (municipalities=321, years=4)	1284	1284	1284	1284

Note: The table shows β coefficients from estimating equation (1) and (2) by OLS and 2SLS. As an illustration, column (3) shows that one standard deviation increase in the share of missing men (in year t) leads to a fall in average literacy rates (in year t) in 51% of a standard deviation. Column (4) shows that one standard deviation increase in the share of missing men (in year t) led to an increase in average literacy rates (10 years later) in 35% of a standard deviation. Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels.

TABLE 5: Long-run effect of missing men on human capital, 1981-2011

	Share of individuals with secondary education			
	1981	1991	2001	2011
	(1)	(2)	(3)	(4)
	Panel A: 2SLS			
Share of missing men (1900-1930)	0.14	0.21*	0.40***	0.46***
	[0.10]	[0.12]	[0.11]	[0.13]
	Panel B: OLS			
Share of missing men (1900-1930)	-0.09**	-0.09*	-0.02	0.04
	[0.05]	[0.05]	[0.05]	[0.06]
Controls	Yes	Yes	Yes	Yes
Kleibergen Paap F-statistic	50.79	50.79	50.79	50.79
Mean dependent variable	0.05	0.24	0.36	0.59
Standard deviation dependent variable	0.03	0.07	0.10	0.11
Mean missing men 1900-1930	0.14	0.14	0.14	0.14
Standard deviation missing men 1900-1930	0.07	0.07	0.07	0.07
Observations	314	314	314	314

Note: The table shows β coefficients of estimating equation (4) by OLS and 2SLS. As an illustration, column (4) in Panel A shows that one standard deviation increase in the average share of missing men (1900-1930) leads to an increase in the share of individuals with completed secondary education (in 2011) in 46% of a standard deviation. Historical controls include population size (1900), male and female literacy rates (1900) and wealth and local taxes (1855) at the municipality level. Clustered standard errors at the district level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 22.

TABLE 6: Explaining the creation of migrant associations and schools, 1910-1960

	At least one association	Number of associations	Number of migrant schools	
Covariates standardized	(1)	(2)	(3)	(4)
Share of missing men (1900-1930)	0.11	1.07	2.25**	1.67**
	[0.14]	[0.72]	[0.90]	[0.68]
Altitude	-0.08*	-0.35*	-0.53**	-0.34*
	[0.04]	[0.19]	[0.23]	[0.18]
Municipality size	0.07	0.90***	0.88**	0.39
	[0.05]	[0.34]	[0.41]	[0.26]
Population settlements	0.07*	0.19	0.19	0.08
	[0.04]	[0.27]	[0.31]	[0.24]
Access to sea	-0.13	0.24	0.54	0.41
	[0.10]	[0.47]	[0.51]	[0.41]
Access to train	0.10	0.47	0.58	0.32
	[0.08]	[0.34]	[0.44]	[0.35]
Distance to port	-0.07	-0.70**	-0.35	0.03
	[0.06]	[0.29]	[0.40]	[0.30]
Distance to capital	0.05	0.09	0.11	0.06
	[0.04]	[0.17]	[0.23]	[0.18]
Population (1860)	0.01	-0.03	-0.21	-0.20
	[0.03]	[0.13]	[0.18]	[0.14]
Men's literacy (1860)	0.01	-0.10	-0.45*	-0.40**
	[0.04]	[0.19]	[0.24]	[0.18]
Women's literacy (1860)	0.01	0.21	0.44*	0.33**
	[0.03]	[0.17]	[0.23]	[0.16]
Wealth per capita (1855)	0.05	0.17	-0.08	-0.17
	[0.03]	[0.16]	[0.22]	[0.17]
Local contributions per capita (1855)	-0.11***	-0.48**	-0.33	-0.07
	[0.03]	[0.19]	[0.24]	[0.18]
Land inequality (1860)	-0.05	-0.04	-0.34	-0.32*
	[0.04]	[0.17]	[0.22]	[0.17]
Rainfall variability (1900-1930)	0.01	-0.44	-1.00***	-0.75***
	[0.07]	[0.33]	[0.35]	[0.28]
Schools per 1000 people (1875)	-0.10***	-0.50***	-0.45**	-0.18
	[0.03]	[0.14]	[0.19]	[0.14]
Children enrollment rates (1875)	0.01	0.30*	0.31	0.14
	[0.03]	[0.18]	[0.21]	[0.16]
Migrants to Rio de la Plata (1900-1930)	0.02	-0.08	-0.27*	-0.23*
	[0.03]	[0.16]	[0.16]	[0.12]
Number of migrant associations				0.54***
				[0.08]
Mean of dependent variable	0.50	1.91	0.92	0.92
Adjusted R-squared	0.18	0.23	0.19	0.48
Observations	321	321	321	321

Note: The table shows the result of estimating equation (8) instrumenting the average share of missing men (1900-1930) by the adult sex ratio in 1860. Covariates are standardized to ease interpretation. As an illustration, column (4) shows that one standard deviation increase in the share of missing men (1900-1930) leads to an increase of 1.67 migrant schools on average. Clustered standard errors at the district level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels.

TABLE 7: Impact of migrants' associations and schools (1910-1960) on education and mobility

	Educational attainment (share of people with completed primary)			Emigration (size of cohort c relative to $c-1$)		
	(1)	(2)	(3)	(4)	(5)	(6)
Opening of a migrant association	-0.001		-0.007*	-0.023		-0.033
	[0.003]		[0.004]	[0.022]		[0.023]
Opening of a migrant school		0.007***	0.01***		0.009	0.017**
		[0.002]	[0.003]		[0.008]	[0.009]
Mean of dependent variable	0.30	0.30	0.30	1.18	1.18	1.18
Standard deviation dependent variable	0.27	0.27	0.27	0.55	0.55	0.55
Mean of explanatory variable	1.04	0.60	-	1.04	0.60	-
Standard deviation explanatory variable	1.67	1.49	-	1.67	1.49	-
Municipality and cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations (municipalities=312, cohorts=8)	2496	2496	2496	2496	2496	2496

Note: The table shows the results of estimating equation (7) by OLS. As an illustration, column (3) shows that the foundation of a new migrant association when individuals are younger than 10, reduces their likelihood of completing primary education by 0.07% on average. Similarly, the opening of a new school financed by migrants increased their chances to complete primary education by 1%. Note that most association financed several school having therefore a larger positive impact on educational attainment. Column (6) shows that the foundation of a new migrant association when individuals are younger than 10, reduces the size of their cohort by 3.3%. Similarly, the opening of a new school financed by migrants increased the size of their cohort by 1.7%. These results suggest that migrants' associations fostered more migration, but if they financed at least two schools, the net impact is zero. All specifications include cohort and municipality fixed effects. The sample comprises all 5-year cohort groups between 1907 and 1947. Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels.

TABLE 8: Long-run effects of missing men on beliefs about education and effort

	Value of education and effort (0-10 scale)			
	(1)	(2)	(3)	(4)
Share of missing men (1900-1930)	0.21	0.23**	0.21	0.22*
	[0.16]	[0.11]	[0.14]	[0.12]
Individual controls	Yes	No	No	Yes
Historical controls	No	Yes	No	Yes
Educational attainment	No	No	Yes	Yes
Mean of dependent variable	4.98	4.98	4.98	4.98
Standard deviation dependent variable	2.33	2.33	2.33	2.33
Mean share of missing men (1900-1930)	0.11	0.11	0.11	0.11
Standard deviation share of missing men (1900-1930)	0.06	0.06	0.06	0.06
Observations	1335	1335	1335	1335

Note: The table shows β coefficients of estimating equation (8) by 2SLS. As an illustration, column (4) shows that one standard deviation increase in the average share of missing men (1900-1930) leads to a 22% of standard deviation increase in reported value for education and effort (e.g., 0.5 points in a 0-10 scale). The average share of missing men (1900-1930) is instrumented by the proxy for migrant networks (i.e., adult sex ratio in 1860). Individual controls include sex, age, and marital status. Historical controls include population size (1900), male and female literacy rates (1900), wealth and local taxes (1855) at the municipality level. Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 28.

Online Appendix

A   Additional Figures and Tables

FIGURE A1: Location of Galicia (Spain)
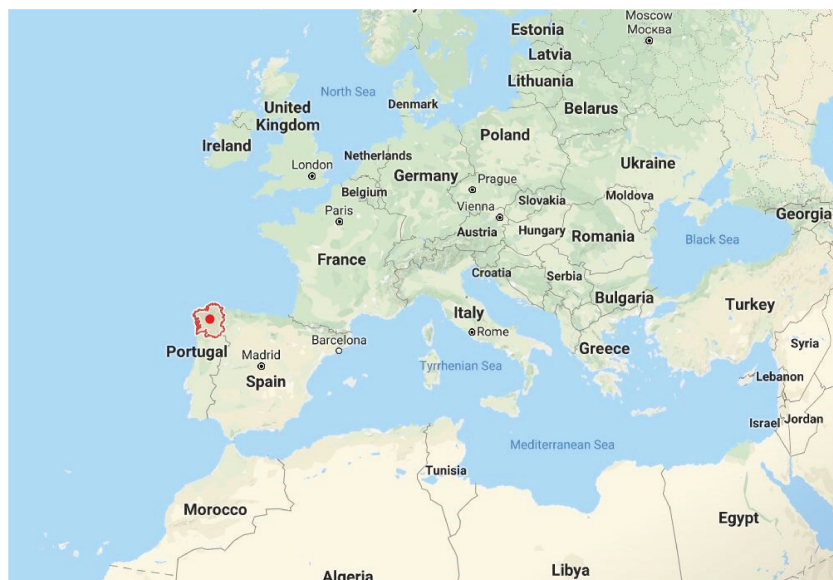
Note: Return to page 8.

FIGURE A2: Yearly departures and returns of Galicians, 1835-1935
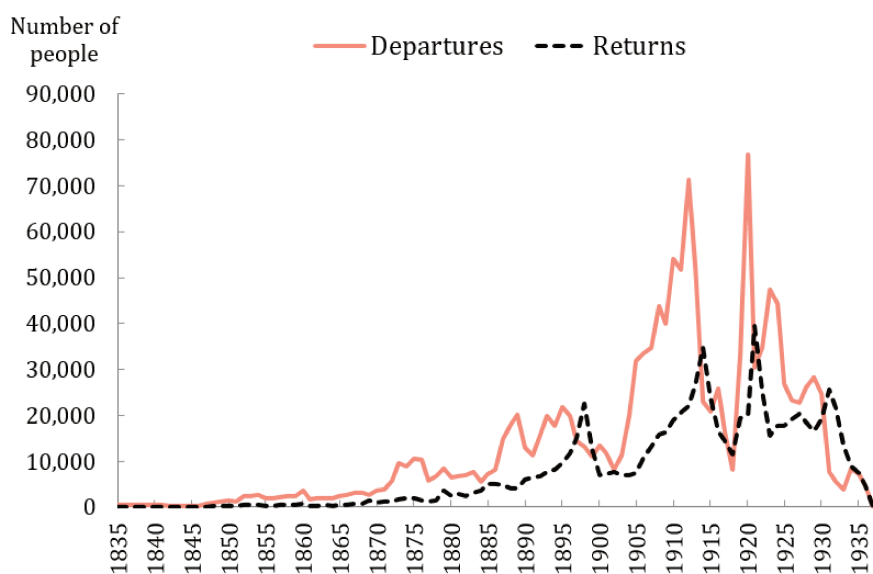
Source: Vázquez-González (1999) and Official Migration Statistics.
Note: The figure displays the number of departures (returns) of Galicians to (from) Latin America every year. These migration flows are probably an underestimation because clandestine migration was not recorded. Return to page 8.

— 90 —

CENSO DE LA POBLACIÓN DE 1900. PROVINCIA DE LA CORUÑA.

AYUNTAMIENTOS.	NÚMERO DE CÉDULAS RECOGIDAS	RESIDENTES				(3) TRANSEUNTES.		(1+3) TOTAL DE LA POBLACIÓN DE HECHO.	(1+2) TOTAL DE LA POBLACIÓN DE DERECHO.
		(1) Presentes.		(2) Ausentes.					
		Varones.	Hembras.	Varones.	Hembras.	Varones.	Hembras.		
1 Abegondo	1 862	3 527	3 841	72	9	5	3	7 376	7 449
2 Ames	2 058	2 687	4 396	910	58	10	4	7 097	8 051
3 Aranga	1 049	2 267	2 441	80	50	2	»	4 710	4 838
4 Ares	1 315	1 619	2 472	483	41	105	13	4 209	4 615
5 Arteijo	2 238	4 658	4 925	366	211	25	21	9 629	10 160
6 Arzúa	2 131	4 145	4 861	154	102	17	13	9 036	9 262
7 Baña (La)	1 687	2 204	3 548	700	41	17	28	5 797	6 493
8 Bergondo	1 410	2 493	3 346	506	55	10	20	5 869	6 400
9 Betanzos	2 437	3 928	4 877	149	42	60	83	8 948	8 996
10 Boimorto	920	1 935	2 098	27	10	5	6	4 044	4 070
11 Boiro	2 229	3 987	4 919	313	74	23	24	8 953	9 293
12 Boqueijón	1 035	1 861	2 352	343	67	11	8	4 232	4 623
13 Brión	1 940	2 080	3 571	894	96	12	18	5 681	6 641
14 Buján	1 107	1 814	2 588	436	62	3	4	4 409	4 900
15 Cabana	1 343	2 111	2 616	75	19	»	»	4 727	4 821
16 Cabañas	884	1 407	1 912	194	13	1	2	3 322	3 526
17 Camariñas	1 242	1 688	2 438	418	28	5	22	4 153	4 572
18 Cambre	1 592	3 181	3 498	109	70	26	17	6 722	6 858
19 Capela	909	1 819	2 141	39	12	»	»	3 960	4 011
20 Carballo	3 385	6 017	6 982	374	181	21	12	13 032	13 554
21 Carnota	1 275	2 427	3 045	181	22	22	13	5 507	5 675
22 Carral	1 403	2 468	2 595	19	14	8	11	5 082	5 096
23 Castro	1 171	2 243	2 677	98	7	2	1	4 923	5 025
24 Cedeira	1 263	2 347	2 872	38	16	»	»	5 219	5 273
25 Cee	1 095	1 618	2 394	338	46	23	25	4 060	4 396
26 Cerceda	1 082	2 271	2 532	39	17	»	»	4 803	4 859
27 Cerdido	793	1 313	1 733	88	17	5	9	3 060	3 151
28 Cesuras	1 231	2 358	2 629	2	5	1	1	4 989	4 994
29 Coirós	661	1 206	1 341	147	95	3	»	2 550	2 789
30 Conjo	1 956	3 202	4 182	506	92	5	9	7 398	7 982
31 Corcubión	460	563	902	172	30	49	37	1 551	1 667
32 Coristanco	1 924	3 005	3 489	162	107	1	»	6 495	6 763
33 CORUÑA (LA)	10 809	18 469	23 858	1 499	231	1 128	516	43 971	44 057
34 Culleredo	1 948	3 631	3 819	21	20	22	23	7 495	7 491
35 Curtis	970	2 047	2 271	28	27	9	8	4 335	4 373
36 Dodro	835	1 263	1 890	205	26	12	16	3 181	3 384
37 Dumbría	912	1 473	2 053	201	25	»	»	3 526	3 752
38 Enfesta	1 004	1 889	2 276	68	4	15	12	4 192	4 237
39 Fene	1 682	2 196	3 180	392	2	6	6	5 388	5 770
40 Ferrol (El)	6 444	11 277	12 988	1 797	195	636	380	25 281	26 257
41 Finisterre	1 389	2 014	2 671	368	51	18	5	4 708	5 104
42 Frades	878	1 539	1 774	57	28	14	1	3 328	3 398
43 Irijoa	853	1 667	1 875	23	38	13	6	3 561	3 603
44 Lage	840	1 305	1 899	212	12	14	27	3 245	3 428

Note: The image shows a typical page in the census chapters. Each row represents a municipality. The population is classified between present (1), missing (2) and transients (3).

FIGURE A4: Emigration and missing men across Spanish provinces, 1920
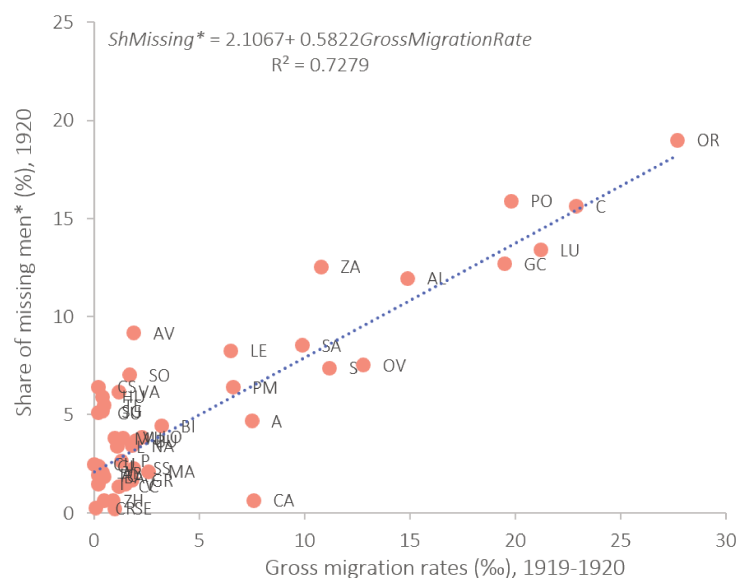
Source: Sánchez-Alonso (1995) and Spanish population census (1920).
Note: "Migration" stands for sea departures. "Missing" stands for residents absent the day of the census. Transients (passers-by) were subtracted from the missing to avoid double counting. Return to page 10.

FIGURE A5: Emigration and missing residents across Italian provinces, 1901
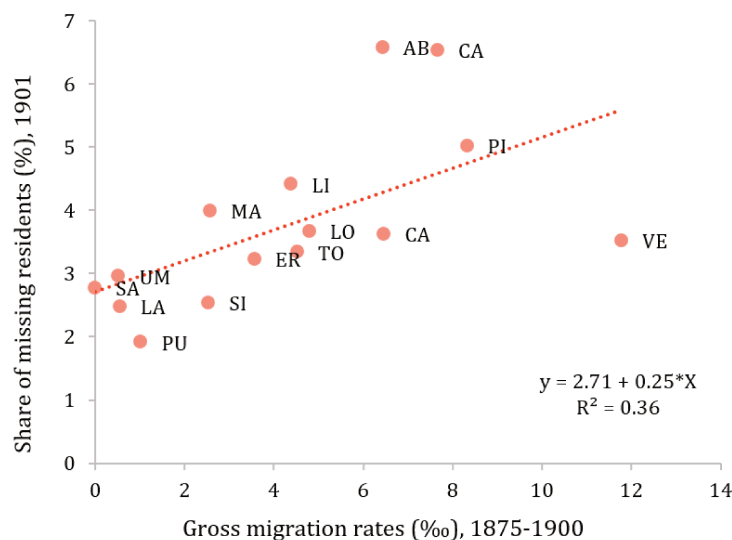
Source: Italian population census (1901) and migration statistics.
Note: PI (Piemonte), LI (Liguria), LO (Lombardia), VE (Veneto), ER (Emilia Ducati/Romagna), TO (Toscana), MA(Marche), UM(Umbria), LA (Lazio), AB (Abruzzi e Molise), CA (Campania), PU (Puglie), BA (Basilicata), CA (Calabrie), SI (Sicilia), SA (Sardegna). Return to page 10.

FIGURE A6: Missing and transient men in Galicia and Spain, 1877-1950
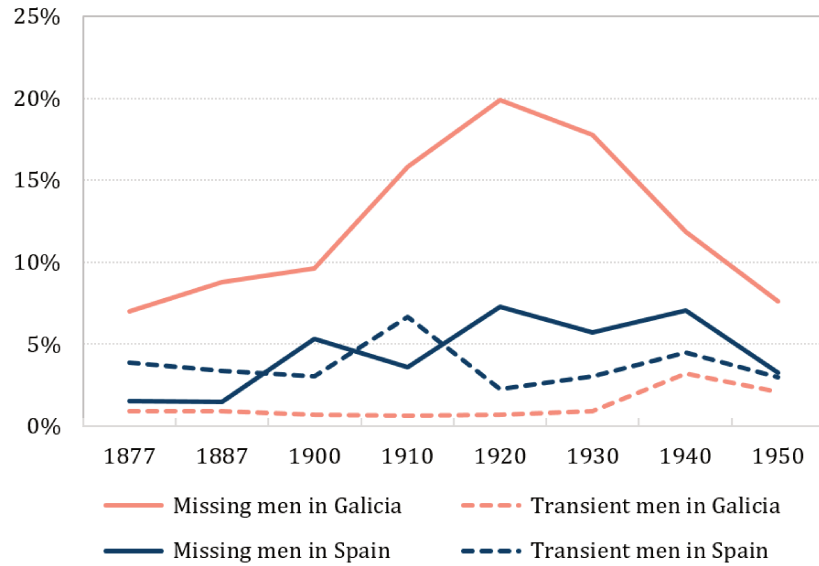
Note: The figure shows the number of missing (i.e., absent the census day) and transient (i.e., in a municipality other than where they regularly reside) men as share of the total male population in Galicia and Spain. Return to page 10.

FIGURE A7: Military draft in Galicia, 1890-1934
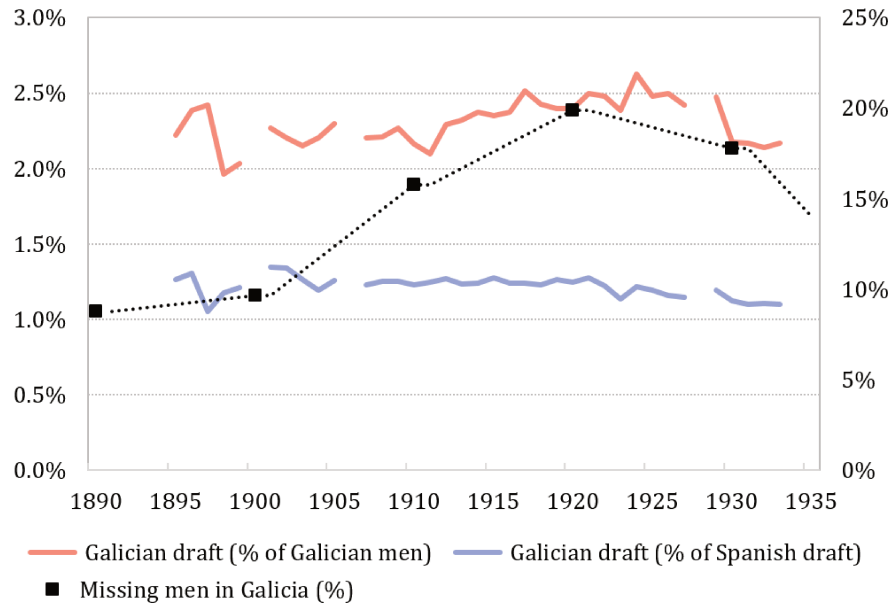
Note: The figure shows the number of Galician men drafted into the military as a share of all Galician men (left axis) and as a share of all Spanish men drafted (right axis). Return to page 10.

F<small>IGURE</small> A8: Share of missing women across Galician municipalities (%), 1877-1930
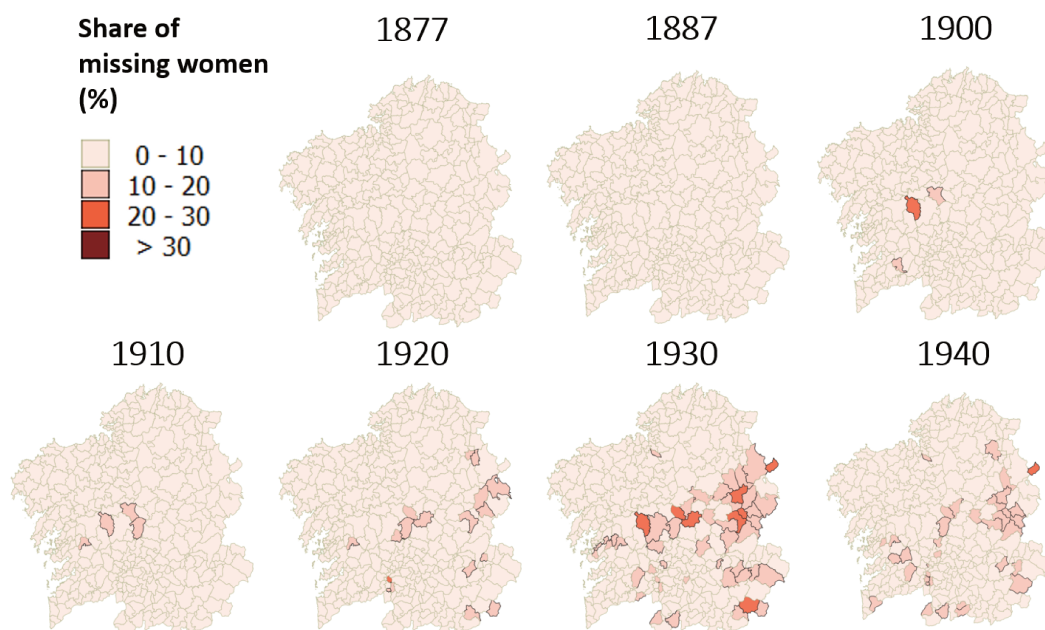
Source: Spanish population census (1877- 1930).
Note: The figure shows the share of women reported as missing in the population censuses of different years. The stronger the intensity of the color the higher the share.

F<small>IGURE</small> A9: Economic growth at destination and origin, 1860-1930
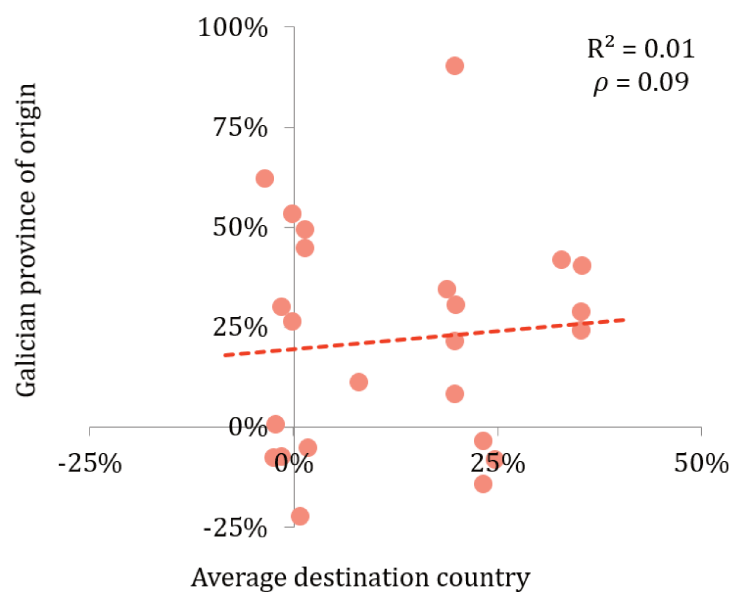
Source: Maddison (2018) for GDP in Latin America, Díez-Minguela et al. (2018) for GDP in Galician provinces, Vázquez-González (1999) for destinations by province.
Note: The x-asis represents the deviations from the GDPpc trend in the average destination country and the y-axis the deviations from the GDPpc trend in the corresponding Galician province. The correlation between both series is 0.09. Return to page 14.

FIGURE A10: Relationship between adult sex ratio (1860) and other covariates
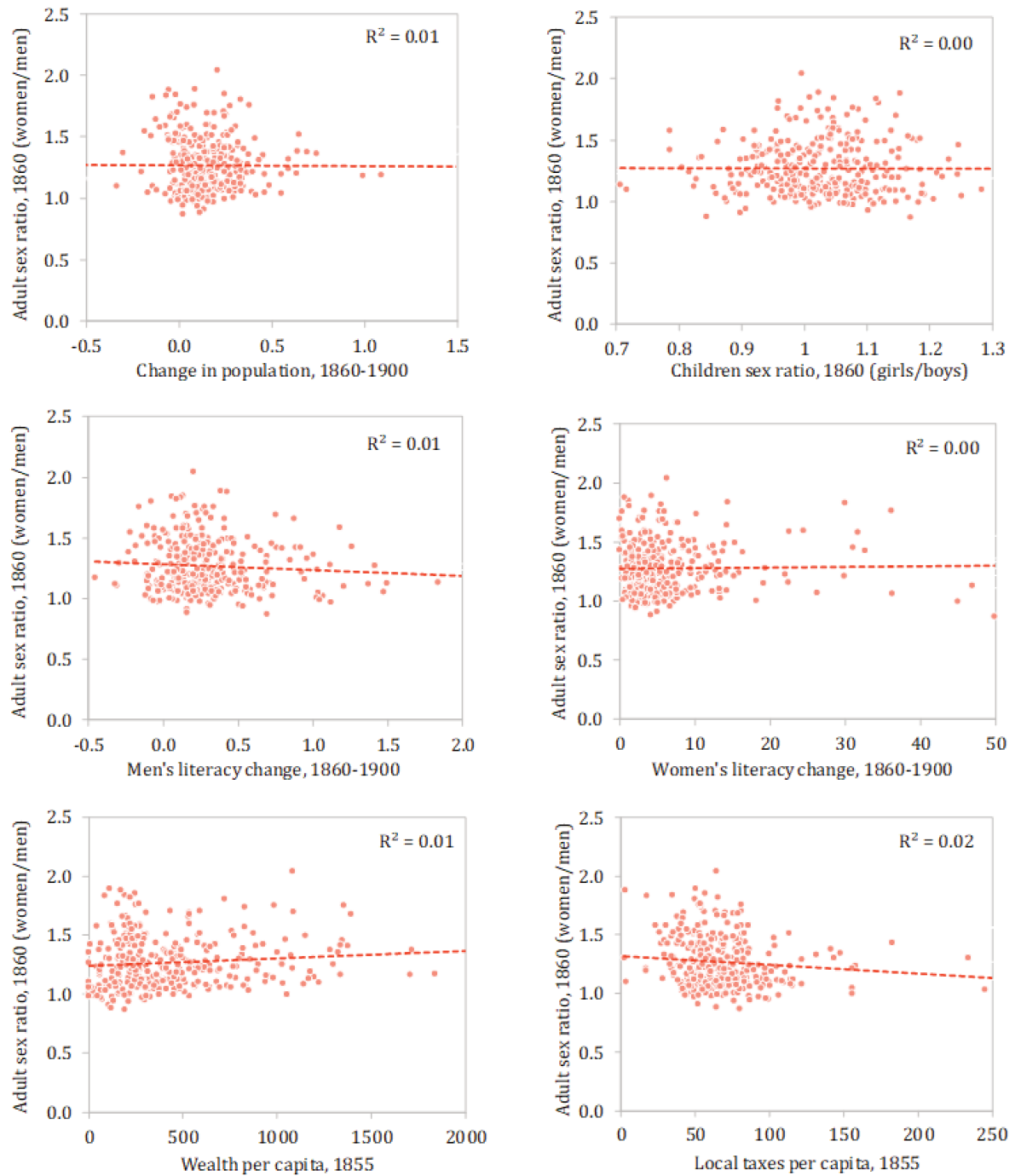
Note: The different figures show the relationship between the adult sex ratio in 1860 (i.e., proxy for pioneer migrant networks) and other covariates measured before the mass emigration era (1900-1930). Return to page 14.

FIGURE A11: Sex ratio in 1860 and average share of missing men in 1900-1930
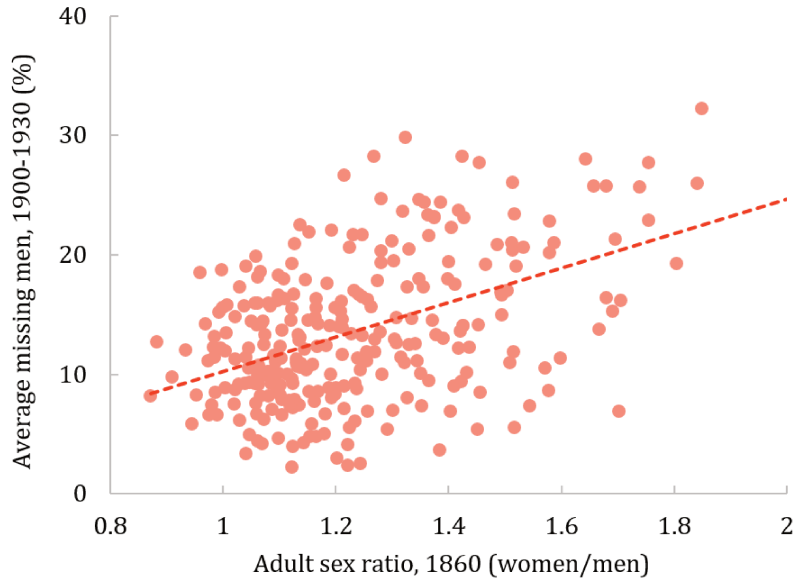
Note: The figure shows the relationship between the adult sex ratio in 1860 (i.e, proxy of pioneer migrant networks) and the average share of missing men between 1900 and 1930 across all municipalities. Return to page 14.

FIGURE A12: Proximity to the port and average share of missing men in 1900-1930
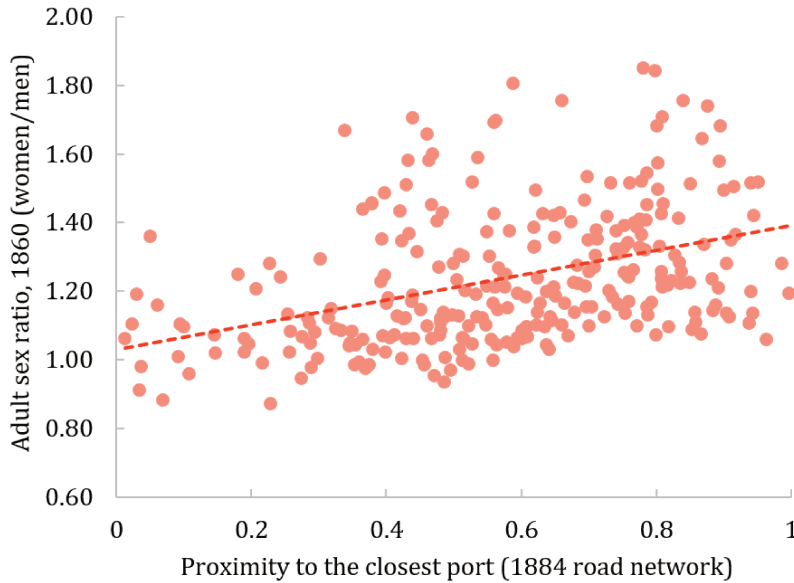
Note: The figure shows the relationship between the walking distance to the closest port and the adult sex ratio in 1860 (i.e, proxy for pioneer migrant networks) across all municipalities. Note that the measure of distance has been normalized to range between 0 and 1. Return to page 14.
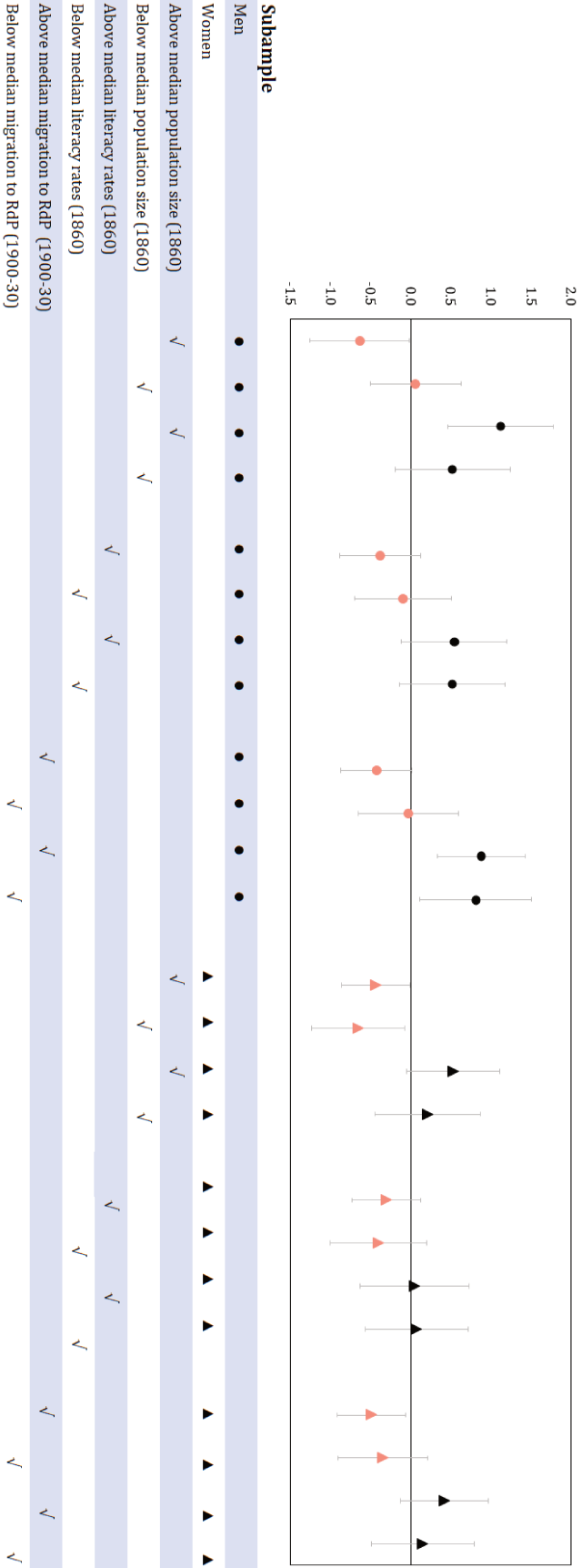
FIGURE A13: Heterogeneity in the short and medium-run effects of missing men

Note: The figure shows the results of estimating equations (1) and (2) for different subsamples. Red circles indicate β coefficient for equation (1), the short-run effect of missing men in literacy rates, while black circles β coefficient for equation (2), the medium run effect of missing men on literacy rates. The share of missing men (in year t) is instrumented with the pull factor (in year t) and its interaction with a proxy for migrant networks. The brackets indicate confidence interval at the 95% level. Return to page 21

125

FIGURE A14: Short and medium-run effects with alternative instruments

Short run	Medium rum

Network									
Adult sex ratio, 1860	●	●	●			●	●	●	
Distance to the closest port				■	■	■			■	■	■
Rainfall variability, 1840-1860					▲	▲	▲			▲	▲	▲
Pull factor									
Municipality weights (1900-30)	√		√		√		√		√		√
District weights (1900-30)		√		√		√		√		√		√
Province weights (1885-95)			√		√		√		√		√		√
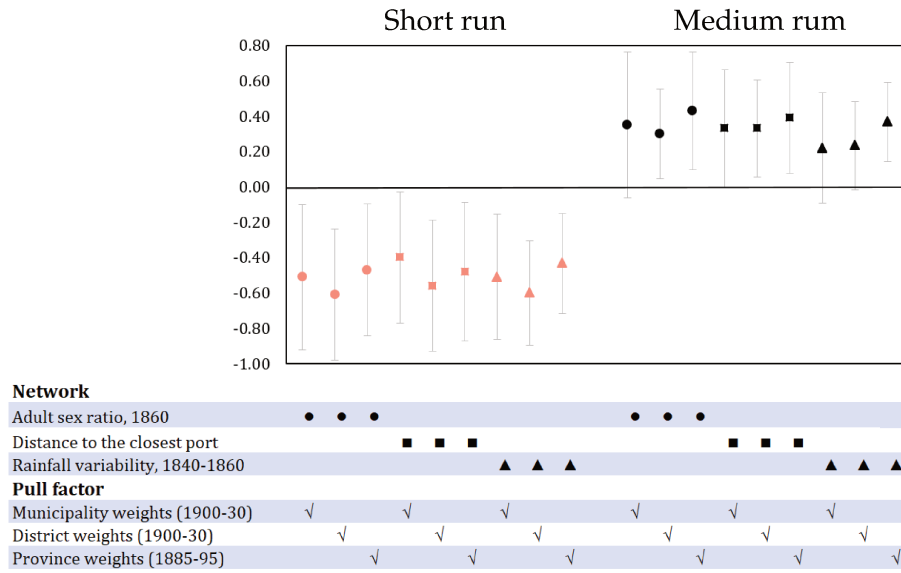
Note: The figure shows the results of estimating equations (1) and (2) with alternative instruments for the pull and network variables. Red circles indicate the β coefficient for equation (1), the short-run effect of missing men in literacy rates, while black circles the β coefficient for equation (2), the medium run effect of missing men on literacy rates. The brackets indicate confidence interval at the 95% level. Return to page .

FIGURE A15: Robustness checks of short and medium-run effects

Short run	Medium rum

Sample exclusion								
Ten km. from a port	x				x			
Extreme population		x				x		
Extreme literacy rates			x				x	
Extreme missing men				x				x
Province of Coruña			x				x	
Province of Lugo				x				x
Province Ourense			x				x	
Province of Pontevedra				x				x
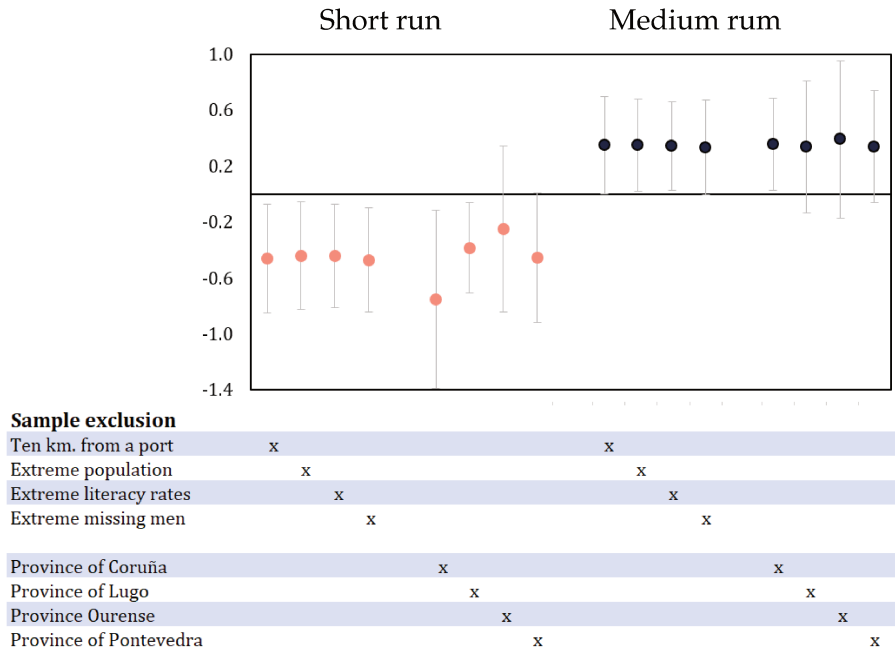
Note: The figure shows the results of estimating equations (1) and (2) with different sub-samples. Red circles indicate β coefficient for equation (1), the short-run effect of missing men in literacy rates, while black circles β coefficient for equation (2), the medium run effect of missing men on literacy rates. The brackets indicate confidence interval at the 95% level. Return to page .
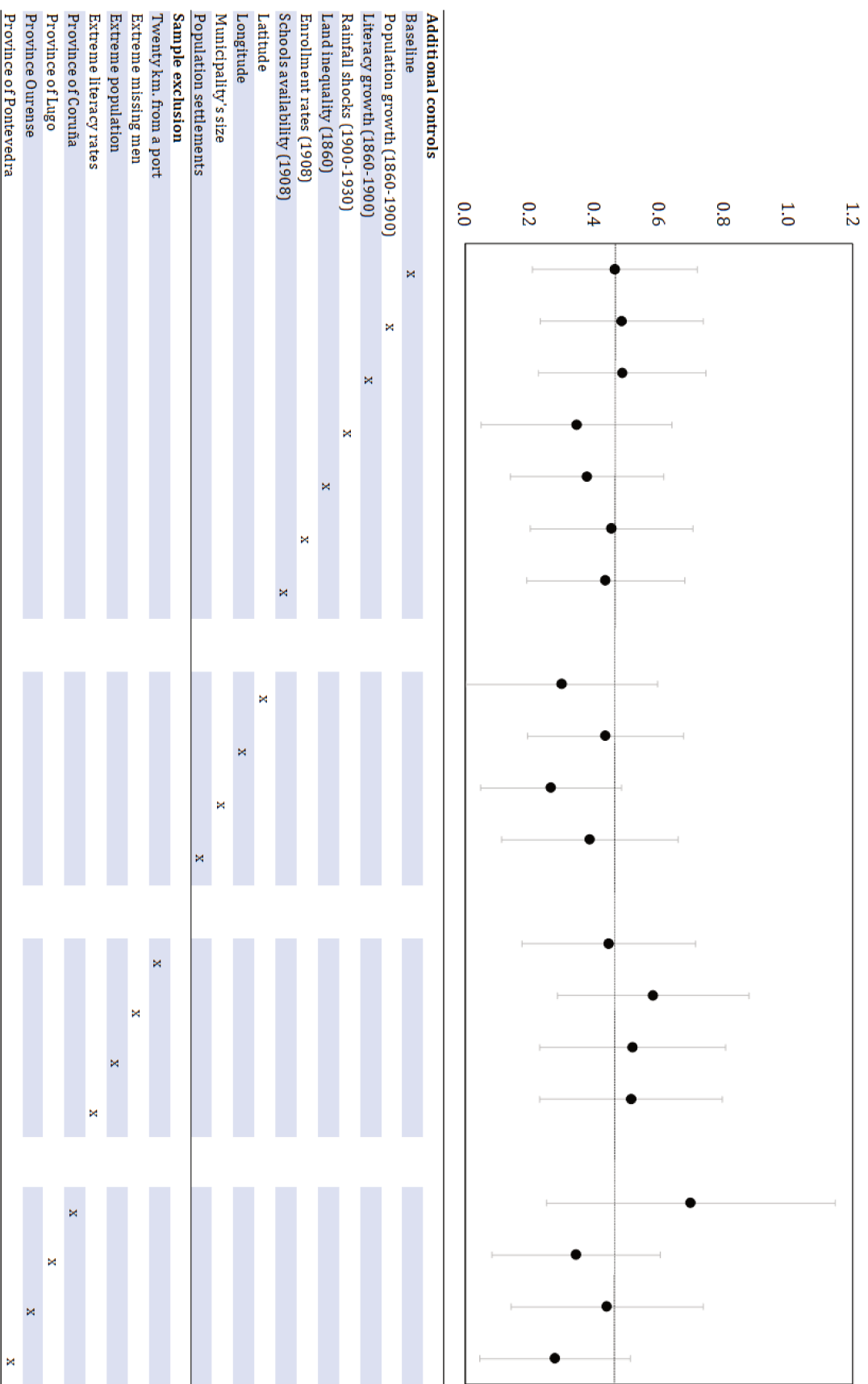
FIGURE A16: Robustness checks of long-run effect of missing men (1900-30) on educational attainment capital (2011)

Note: The figure shows the results of estimating equation (4) by 2SLS with different controls and sub-samples . The dependent variable is the share of individuals with completed secondary education in 2011 and the main explanatory variable the average share of missing men (1900-1930). The point estimates are β coefficients and the horizontal line represent the baseline estimate. The average share of missing men (1900-1930) is instrumented with the proxy for migrant networks (adult sex ratio in 1860). Baseline controls include population (1860), literacy rates (1860), wealth per capita (1855) and local contributions per capita (1855) at municipality level. The brackets indicate confidence interval at the 95% level based on clustered standard errors at the district level. Return to page 23.

127

FIGURE A17: Evolution of the size of the cohort born in 1885-1899
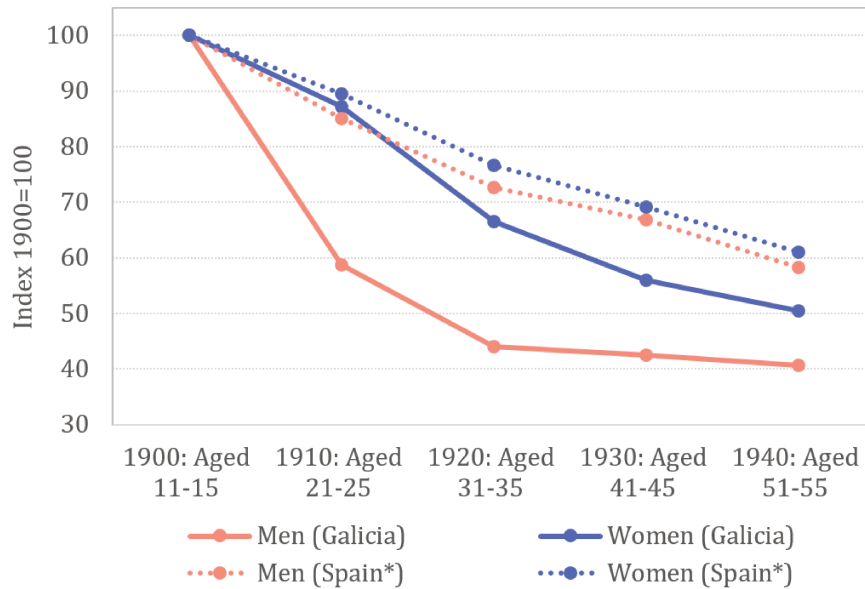
Note: The figure shows the evolution of the cohort of individuals born in 1885-1889. The size of the cohort has been normalized to 100 in 1900. For instance, the figure shows that the size of this cohort of Galician men decreased by 40% between 1900 and 1910. The figures for Spain do not include Galicia. Return to page 25.

FIGURE A18: Evolution of literacy rates of the cohort born in 1885-1889
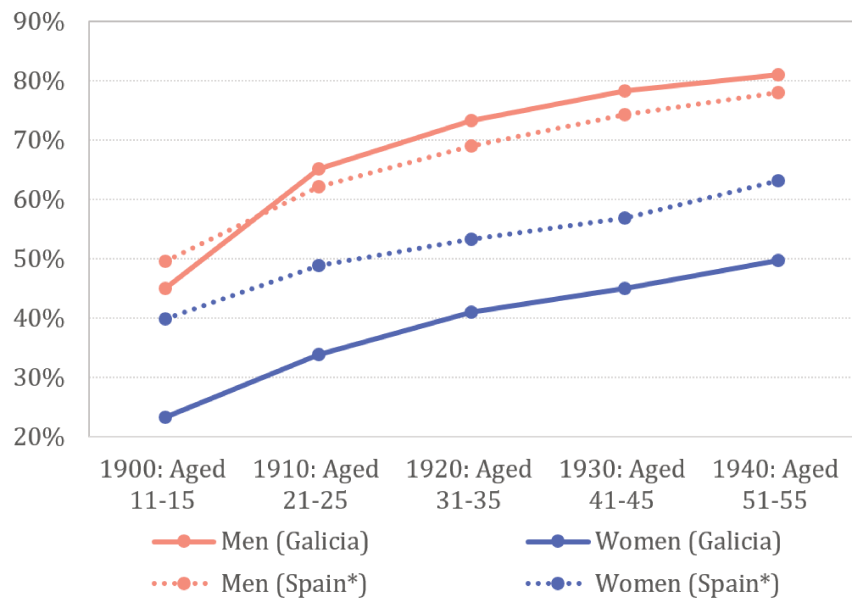
Note: The figure shows the evolution of literacy rates for the cohort of individuals born in 1885-1889. For instance, the figure shows that the literacy rate of this cohort of Galician men increased from 45% to 66% between 1900 and 1910. The figures for Spain do not include Galicia. Return to page 25.

FIGURE A19: Missing men and cohort size changes, 1900-1910
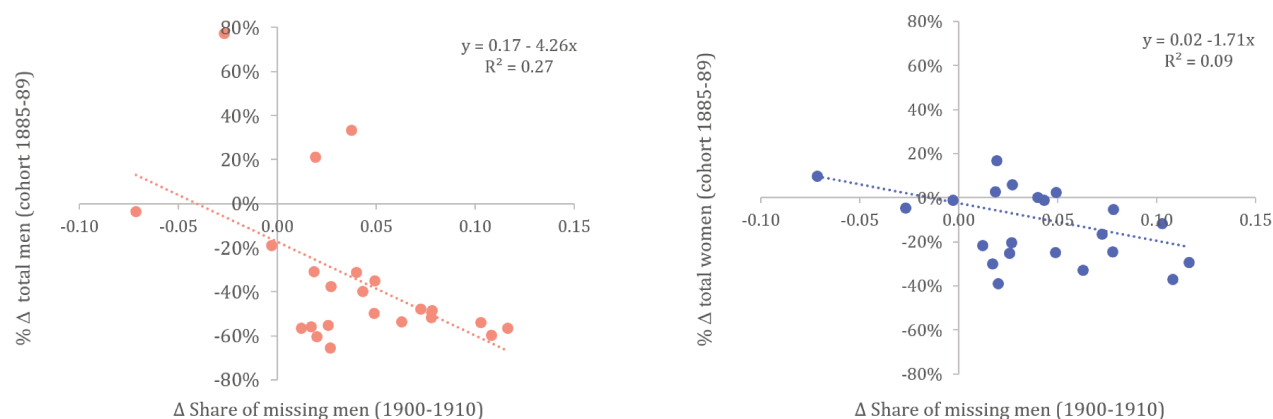
Note: The figure shows the relationship between changes in the share of missing men (1900-10) and changes in the size of the cohort born in 1885-1889 (1900-10) across the 23 largest Galician districts. For example, a value of -50% means that the size of the cohort halved between 1900 and 1910. Return to page 21.

FIGURE A20: Missing men and literacy rates changes, 1900-1910
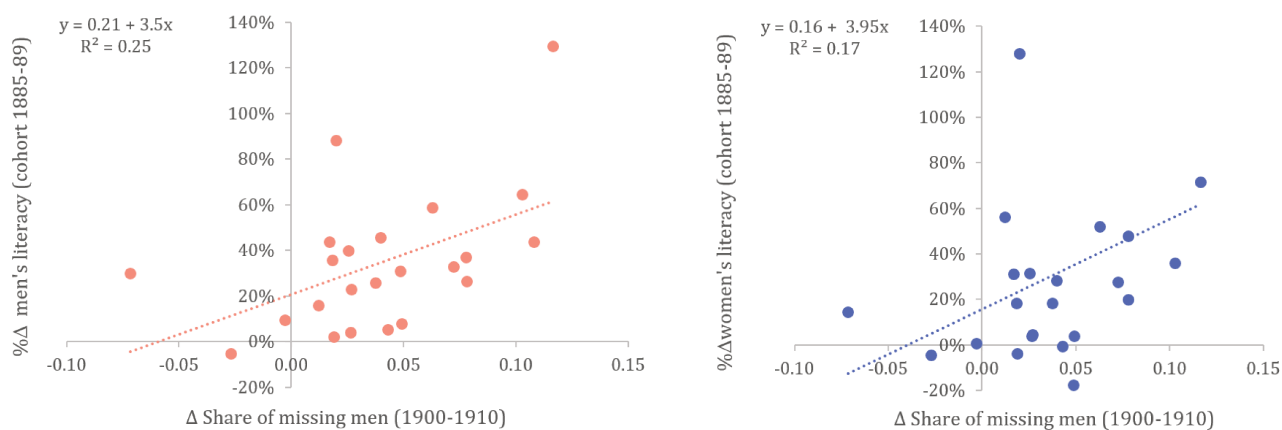
Note: The figure shows the relationship between changes in the share of missing men (1900-10) and changes in literacy rates of the cohort born in 1885-1889 (1900-10) across the 23 largest Galician districts. For instance, a value of 20% means that the average literacy rate of the cohort increased by 20% between 1900 and 1910. Return to page 21.

FIGURE A21: Examples of schools financed by Galician migrants in their hometowns

Escolas Alianza Aresana, Ares, 1909

Escola La Devesana, Ortigueira, 1912

Escola de Cecebre, Cambre, 1922

Escola de Covas, Negreira, 1922

Grupo escolar de Piñeiro, Cedeira, 1924

Escola Luz de Riberas del Sor, Mañón, 1917

Escola da Torre, Cabanas, 1929

Escola Rosalía de Castro, Rianxo, 1929

Casa-escola Marcial González, Estorde, Cee, 1931

Casa-escola de Tallós, Dodro, c. 1930

Casa-escola de Seoane, A Baña, 1932

Escola da Avispeira, A Barqueira, Cerdido, 1932

Note: Pictures extracted from the project "Luces de Alén Mar". Information of around 300 migrants' schools is available here: http://mapas.consellodacultura.gal/escolas/. Return to page 25.

FIGURE A22: Foundation of migrants' associations and schools, 1900-1960
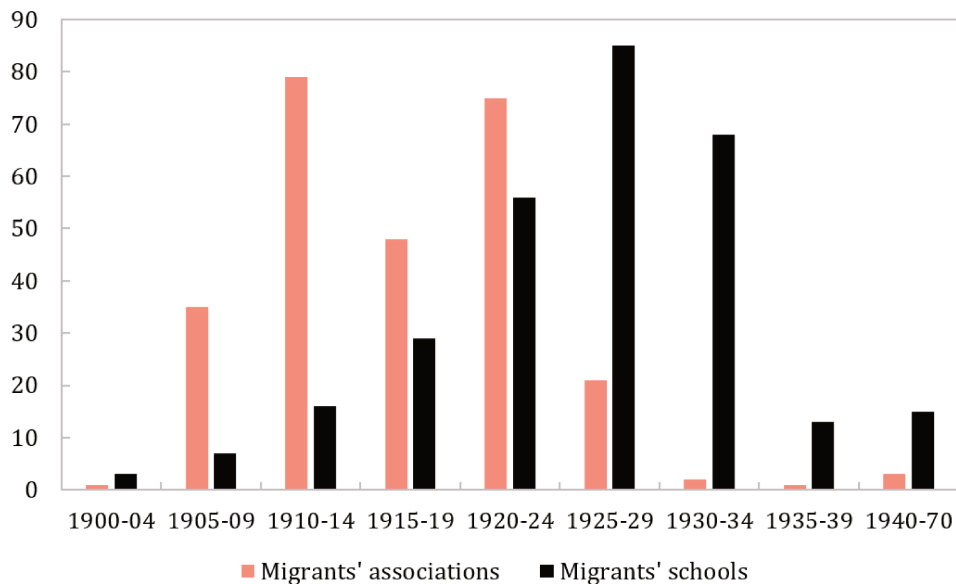
Note: The figure shows the number of Galician migrants' associations (with an educational goal) and the new schools financed by migrants in 5-year intervals. Return to page 25.

FIGURE A23: Location of migrants' schools over time, 1900-1960
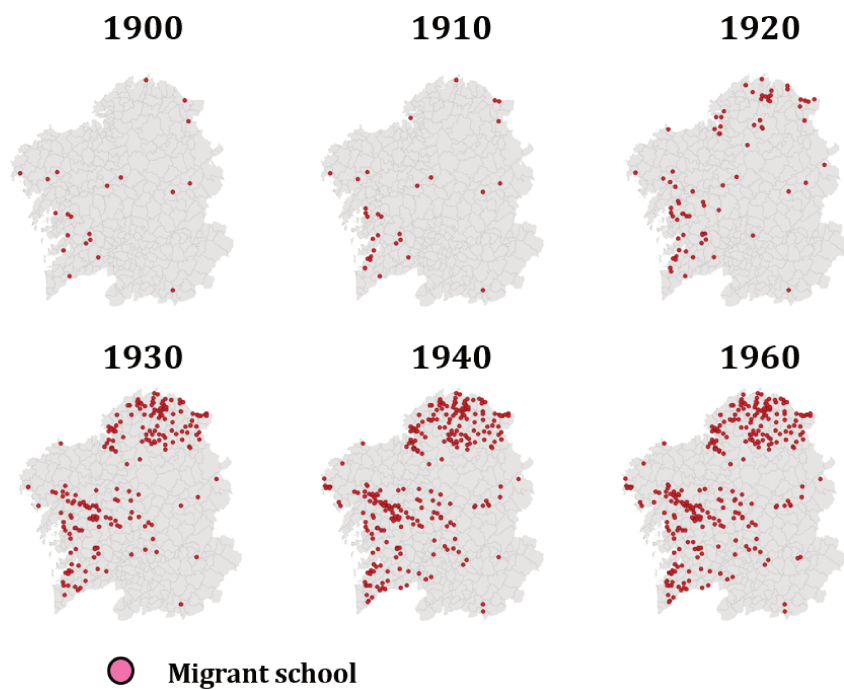
Note: The figure shows the location of all schools financed by migrants at different points in time. Return to page 25.

FIGURE A24: Location of migrants' schools and intensity of missing men (1900-1930)
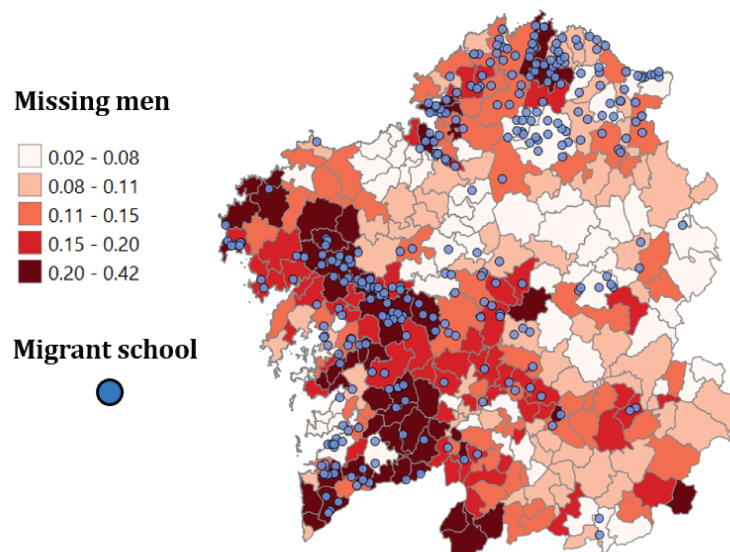
Note: The figure shows the average share of missing men between 1900 and 1930 and the location of all the migrants' schools founded in Galicia until 1960. Return to page 25.

Tables

TABLE A1: Characteristics of Galician migrants

| | **Sex, age, civil status, literacy** | | | **Occupation** | |
	1840-1900	1900-1930		1890-1900	1900-1930
Share of men	81%	67%	Share of farmers	94%	90%
Share younger than 10	7%	6%	Share of merchants	5%	6%
Share aged 10-15	25%	7%	Other occupations	1%	4%
Share aged 15-25	29%	40%		**Literacy**	
Share older than 25	39%	46%		1918	1925-1930
Share of single	65%	70%	Literacy of emigrants	83%	84%
Share of literate	51%	82%	Literacy of returnees	-	82%

Source: Left panel [Sex, age, civil, status, literacy] comes from Vázquez-González (1999). He compiled information of around 20 thousand migrants. Right panel [Occupation/Literacy], comes from the Official Emigration Statistics (see data sources in the Appendix). The data refers to Galician ports. *Note*: Return to page .

TABLE A2: Literacy rates in Galicia across cohorts and time

| | **Men** | | | | | **Women** | | | | |
Ages	1900	1910	1920	1930	1940	1900	1910	1920	1930	1940
11-20	49%	57%	64%	77%	89%	24%	38%	51%	70%	85%
21-30	57%	65%	74%	81%	92%	20%	32%	47%	62%	81%
31-40	59%	62%	71%	79%	87%	16%	24%	37%	52%	69%
41-50	57%	62%	68%	76%	64%	13%	20%	28%	41%	60%
51-60	52%	58%	65%	70%	78%	9%	15%	21%	30%	47%
61-70	48%	52%	61%	66%	73%	8%	12%	17%	22%	34%
71-80	49%	51%	56%	63%	68%	9%	13%	14%	20%	29%
Average	54%	59%	67%	76%	86%	16%	26%	37%	50%	67%
Standard deviation	8%	9%	9%	12%	11%	6%	10%	15%	21%	24%

Source: Spanish population census (1900-1940). *Note*: The table shows the literacy rates of men and women in Galicia in different years and for different age groups. By following cells of a same color diagonally one can look at the evolution of literacy rates of a given cohort over time. Return to page .

TABLE A3: OLS regressions with different instruments and various years

	Share of missing men							
	1877	1887	1900	1910	1920	1930	1877-1930	1900-1930
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Adult sex ratio, 1860	0.34***	0.40***	0.30***	0.24***	0.17***	0.24***	0.26***	0.19***
(women/men)	[0.04]	[0.04]	[0.04]	[0.03]	[0.04]	[0.04]	[0.00]	[0.00]
Adjusted R-squared	0.52	0.51	0.30	0.18	0.11	0.19	0.46	0.32
Rainfall variability, 1830-1860	0.09***	0.13***	0.10***	0.08***	0.24	0.06***	0.46***	0.33***
	[0.02]	[0.03]	[0.03]	[0.03]	[0.02]	[0.03]	[0.00]	[0.00]
Adjusted R-squared	0.17	0.21	0.16	0.10	0.05	0.07	0.35	0.24
Distance to closest port	0.09***	0.13***	0.10***	0.08***	0.24	0.06***	0.46***	0.33***
	[0.02]	[0.03]	[0.03]	[0.03]	[0.02]	[0.03]	[0.00]	[0.00]
Adjusted R-squared	0.10	0.14	0.08	0.04	0.00	0.02	0.30	0.19
Average growth at destination (municipality level)	0.40***	-0.24	0.26***	0.10	0.06	0.06	0.04***	0.07***
	[0.13]	[0.19]	[0.08]	[0.10]	[0.06]	[0.06]	[0.02]	[0.02]
Adjusted R-squared	0.04	0.02	0.04	0.00	0.00	0.00	0.75	0.74
Average growth at destination (district level)	0.67***	-0.54	0.50***	0.09	0.04	0.18***	0.07***	0.12***
	[0.19]	[0.40]	[0.13]	[0.17]	[0.09]	[0.10]	[0.02]	[0.02]
Adjusted R-squared	0.06	0.04	0.09	0.00	0.00	0.03	0.75	0.75
Average growth at destination (province level)	1.04***	-0.42***	0.85***	-0.13	0.07	0.25***	0.06***	0.10***
	[0.27]	[0.17]	[0.18]	[0.15]	[0.13]	[0.08]	[0.02]	[0.03]
Adjusted R-squared	0.11	0.06	0.18	0	0	0.06	0.75	0.74
Municipality ans year FE	No	No	No	No	No	No	Yes*	Yes*
Observations	322	322	322	322	322	322	1932	1288

Note: The table shows β coefficients of regressing the share of missing men (in year t) on different covariates. In columns (7) and (8), I include municipality and year fixed effects depending on the level of variation of the variable of interest. Clustered standard errors at the district (columns 1-6) and municipality (columns 7-8) level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 19.

TABLE A4: Relationship between adult sex ratio in 1860 and other covariates

	Adult sex ratio (women/men, 1860)			
	No controls		Within province	
	β Coefficient	p-value	β Coefficient	p-value
	(1)	(2)	(3)	(4)
Altitude	-0.51	0.00	-0.26	0.00
Access to sea	0.38	0.00	0.16	0.00
Access to road	-0.28	0.00	-0.18	0.00
Access to train	-0.02	0.78	-0.07	0.13
Distance to capital	-0.04	0.52	0.02	0.72
Distance to port	-0.51	0.00	-0.15	0.06
Municipality size	-0.18	0.00	-0.14	0.01
Number of parishes	-0.20	0.00	-0.19	0.00
Population (1860)	0.18	0.00	0.02	0.61
Share of children (1860)	-0.19	0.00	-0.21	0.00
Men's literacy (1860)	0.18	0.00	0.06	0.22
Women's literacy (1860)	0.13	0.02	0.07	0.14
Wealth per capita (1855)	0.03	0.62	-0.02	0.80
Local taxes per capita (1855)	-0.13	0.02	-0.13	0.01
School enrollment rates (1875)	0.03	0.64	0.03	0.62
Schools per thousand people (1875)	0.10	0.17	0.11	0.05
Population change (1860-1900)	0.02	0.72	-0.04	0.34
Men's literacy change (1860-1900)	0.05	0.37	0.09	0.04
Women's literacy change (1860-1900)	0.04	0.52	0.05	0.28
Observations	321		321	

Note: The table shows β coefficients and p-values of multiple regressions of the adult sex ratio (1860) on covariates. As an illustration, the first row and column shows that an increase in the municipality's average altitude in one standard deviation is associated with a 51% standard-deviation decrease in the adult sex ratio (women/men) in 1860. Columns (3) and (4) show the results of similar regressions but including province fixed effects. Return to page 21.

TABLE A5: Impact of missing men on literacy rates by main destination, 1900-1930

	Literacy rates (t)							
	Men				Women			
	High RdP (1900-30)		Low RdP (1900-30)		High RdP (1900-30)		Low RdP (1900-30)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of missing men (t)	-0.38		-0.10		-0.31		-0.40	
	[0.26]		[0.31]		[0.22]		[0.31]	
Share of missing men (10 years before)		0.54*		0.52		0.05		0.07
		[0.34]		[0.34]		[0.35]		[0.33]
Municipality and year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent variable	0.46	0.46	0.44	0.44	0.21	0.21	0.21	0.21
Standard deviation dependent variable	0.12	0.12	0.13	0.13	0.15	0.15	0.16	0.16
Mean share of missing men	0.13	0.13	0.12	0.12	0.13	0.13	0.12	0.12
Satandard deviation share of missing men	0.09	0.09	0.08	0.08	0.09	0.09	0.08	0.08
Observations (municipalities=321, years=4)	644	644	640	640	644	644	640	640

Note: The table shows β coefficients from estimating equations (1) and (2) by 2SLS. The share of missing men (in year *t* is instrumented by the economic growth at destination (in the decade preceding *t*) and its interaction with the proxy for pioneer migrant networks (adult sex ratio in 1860). Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 21.

TABLE A6: Impact of missing men on literacy rates by initial population, 1900-1930

	Literacy rates (t)							
	Men				Women			
	High population 1860		Low population 1860		High population 1860		Low population 1860	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of missing men (t)	-0.43**		-0.03		-0.49**		-0.35	
	[0.23]		[0.32]		[0.22]		[0.29]	
Share of missing men (10 years before)		0.88***		0.81**		0.42		0.15
		[0.28]		[0.36]		[0.28]		[0.33]
Municipality and year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent variable	0.47	0.47	0.43	0.43	0.20	0.20	0.16	0.16
Standard deviation dependent variable	0.12	0.12	0.13	0.13	0.16	0.16	0.15	0.15
Mean share of missing men	0.13	0.11	0.13	0.10	0.13	0.11	0.13	0.10
Satandard deviation share of missing men	0.09	0.08	0.09	0.08	0.09	0.08	0.09	0.08
Observations (municipalities=321, years=4)	644	644	640	640	644	644	640	640

Note: The table shows β coefficients from estimating equations (1) and (2) by 2SLS. The share of missing men (in year *t* is instrumented by the economic growth at destination (in the decade preceding *t*) and its interaction with the proxy for pioneer migrant networks (adult sex ratio in 1860). Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 21.

TABLE A7: Impact of missing men on literacy rates by initial literacy, 1900-1930

	Literacy rates (t)							
	Men				Women			
	High literacy 1860		Low literacy 1860		High literacy 1860		Low literacy 1860	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of missing men (t)	-0.64**		0.06		-0.44**		-0.66**	
	[0.32]		[0.29]		[0.22]		[0.30]	
Share of missing men (10 years before)		1.12***		0.52		0.53*		0.21
		[0.34]		[0.37]		[0.30]		[0.34]
Municipality and year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent variable	0.51	0.51	0.40	0.40	0.26	0.26	0.17	0.17
Standard deviation dependent variable	0.11	0.11	0.12	0.12	0.16	0.16	0.14	0.14
Mean share of missing men	0.14	0.12	0.12	0.10	0.14	0.12	0.12	0.10
Satandard deviation share of missing men	0.09	0.09	0.08	0.08	0.09	0.08	0.09	0.08
Observations (municipalities=321, years=4)	644	644	640	640	644	644	640	640

Note: The table shows β coefficients from estimating equations (1) and (2) by 2SLS. The share of missing men (in year t is instrumented by the economic growth at destination (in the decade preceding t) and its interaction with the proxy for pioneer migrant networks (adult sex ratio in 1860). Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 21.

TABLE A8: Impact of missing men on literacy rates controlling for economic variables

| | Men's literacy rates (t) | | | | | |
	(1)	(2)	(3)	(4)	(5)	(6)
Share of missing men (t)	-0.11		-0.03		-0.22	
	[0.22]		[0.24]		[0.28]	
Share of missing men (10 years before)		0.57***		0.86***		0.64***
		[0.19]		[0.31]		[0.21]
Municipality and year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent variable	0.48	0.48	0.48	0.48	0.48	0.48
Standard deviation dependent variable	0.12	0.12	0.12	0.12	0.12	0.12
Mean share of missing men	0.14	0.11	0.14	0.11	0.14	0.11
Standard deviation share of missing men	0.08	0.09	0.08	0.09	0.08	0.09
Observations (municipalities=321, years=4)	1284	1284	1284	1284	1284	1284

Note: Column (1) controls for GDP per capita in the province and its interaction with the distance to the capital. Column (2) controls for rainfall variability during the precedent decade. Column (3) controls for the size of exports and imports to Latin America and its interaction with the distance to the closest port. Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 21.

TABLE A9: Impact of missing population and missing women on literacy rates

	Literacy rates (t)							
	All		Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of missing population (t)	-0.53**		-0.31		-0.58**			
	[0.25]		[0.24]		[0.28]			
Share of missing population (10 years before)		0.33**		0.62***		0.13		
		[0.17]		[0.18]		[0.19]		
Share of missing women (t)							-0.19	
							[0.37]	
Share of missing women (10 years before)								0.18
								[0.27]
Municipality and year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
KP F -statistic	5.33	14.81	5.33	14.81	5.33	14.81	1.28	4.35
Mean dependent variable	0.35	0.35	0.48	0.48	0.25	0.25	0.25	0.25
Standard deviation dependent variable	0.13	0.13	0.12	0.12	0.15	0.15	0.15	0.15
Mean share of missing	0.09	0.07	0.09	0.07	0.09	0.07	0.04	0.02
Standard deviation share of missing	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.03
Observations (n=321, t=4)	1284	1284	1284	1284	1284	1284	1284	1284

Note: The table shows β coefficients from estimating equations (1) and (2) by 2SLS but using the share of missing population or women as main explanatory variables. The share of missing population or women in year t is instrumented by the economic growth at destination (in the decade preceding t) and its interaction with the proxy for pioneer migrant networks (adult sex ratio in 1860). Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 21.

TABLE A10: Long-run impact of missing men on human capital by birth cohorts, 1930s-1970s

| | Birth cohorts | | | | |
| | 1930s | 1940s | 1950s | 1960s | 1970s |
	(1)	(2)	(3)	(4)	(5)
	Panel A: Completed primary education (%)				
Average share of missing men (1900-1930)	0.35***	0.37***	0.35***	0.31***	0.29**
	[0.13]	[0.12]	[0.12]	[0.13]	[0.14]
Mean dependent variable	0.54	0.73	0.87	0.95	0.98
Standard deviation dependent variable	0.21	0.18	0.12	0.06	0.03
	Panel B: Completed secondary education (%)				
Average share of missing men (1900-1930)	0.30***	0.21***	0.03	-0.04	0.04
	[0.11]	[0.11]	[0.12]	[0.13]	[0.14]
Mean dependent variable	0.10	0.25	0.49	0.75	0.88
Standard deviation dependent variable	0.08	0.14	0.15	0.12	0.08
	Panel C: Completed tertiary education (%)				
Average share of missing men (1900-1930)	0.17	0.21*	0.21*	-0.00	-0.26***
	[0.12]	[0.13]	[0.13]	[0.13]	[0.12]
Mean dependent variable	0.02	0.04	0.07	0.12	0.14
Standard deviation dependent variable	0.02	0.03	0.05	0.06	0.05
Controls	Yes	Yes	Yes	Yes	Yes
KP F-statistic	50.79	50.79	50.79	50.79	50.79
Mean average share of missing men (1900-30)	0.14	0.14	0.14	0.14	0.14
Standard deviation average missing men (1900-30)	0.07	0.07	0.07	0.07	0.07
Observations	315	315	315	315	315

Note: The table shows β coefficients of estimating equation (5) by 2SLS using as dependent variable different measures of human capital by birth cohort (measured in 2001). The average share of missing men (1900-1930) is instrumented by the adult sex ratio in 1860. Historical controls include population size (1900), male and female literacy rates (1900) and wealth and local taxes (1855) at the municipality level. Clustered standard errors at the district level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels.

TABLE A11: Long-run effect of missing men (1900-1930) on various modern outcomes (2018)

	Public schools per 1000 people	Private schools per 1000 people	Public libraries per 1000 people	Local associations per 1000 people	Educational associations (%)
	(1)	(2)	(3)	(4)	(5)
Average share of missing men (1900-1930)	0.02	0.01	0.05	-0.35***	0.44***
	[0.03]	[0.01]	[0.11]	[0.10]	[0.13]
Controls	Yes	Yes	Yes	Yes	Yes
KP F-statistic	50.79	50.79	50.79	50.79	50.79
Mean dependent variable	0.64	0.03	0.27	13.39	0.11
Standard deviation dependent variable	1.10	0.18	0.30	16.60	0.08
Mean missing men 1900-1930	0.14	0.14	0.14	0.14	0.14
Standard deviation missing men 1900-1930	0.07	0.07	0.07	0.07	0.07
Observations	315	315	315	315	315

Note: The table shows β coefficients of different cross-sectional regressions. The average share of missing men (1900-1930) is instrumented by the adult sex ratio in 1860. Historical controls include population size (1900), male and female literacy rates (1900) and wealth and local taxes (1855) at the municipality level. Clustered standard errors at the district level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 23.

TABLE A12: Long-run impact of missing men on beliefs about taxation and public goods (2007-2018)

	Satisfied with public education	Too little spend on public education	More taxes for better public goods (0-10)	We benefit little from taxes	We pay little taxes
	(1)	(2)	(3)	(4)	(5)
Average share of missing men (1900-1930)	0.00	-0.08*	0.02	0.00	-0.08
	[0.01]	[0.05]	[0.08]	[0.05]	[0.06]
Individual controls	Yes	Yes	Yes	Yes	Yes
Historical controls	Yes	Yes	Yes	Yes	Yes
Educational attainment	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable	0.45	0.55	4.98	4.98	0.61
Standard deviation dependent variable	0.50	0.50	2.33	0.60	0.49
Mean share of missing men (1900-1930)	0.11	0.11	0.11	0.11	0.11
Standard deviation share of missing men (1900-1930)	0.06	0.06	0.06	0.06	0.06
Observations	1862	1834	1264	1918	2066

Note: The table shows β coefficients of different cross-sectional regressions. The average share of missing men (1900-1930) is instrumented by the adult sex ratio in 1860. Historical controls include population size (1900), male and female literacy rates (1900) and wealth and local taxes (1855) at the municipality level. Clustered standard errors at the district level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Clustered standard errors at the municipality level in brackets. *, ** and *** indicate significant at the 10, 5 and 1% levels. Return to page 28.

B Summary of data sources

TABLE B1: Main historical sources

Description	Source
Main demographic data (e.g., literate, missing) 1860, 1877, 1887, 1900, 1910, 1920, 1930	Spanish population censuses http://www.ine.es/inebaseweb/libros.do?tntp=71807&L=1
Migrants' associations and schools 1900-1960	Peña Saavedra (1991), Nuñhez Seixas (1992), Naranjo Orovio (1987), Costa Rico (1989), Llorden Miñambres (1996), Díaz Sal (1975)
Galician migrant destinations 1900-1930	Vázquez-González (1991): desertion records, embarkation lists and notary documents.
Pre-migration data (e.g., wealth, taxes) 1855	Diccionario geográfico estadístico histórico de España y sus posesiones de Ultramar https://archive.org/details/diccionariogeogr10madouoft/page/n4
Educational statistics 1874, 1908, 1974	Estadística de la matrícula escolar del año académico 1874-1875, Censo escolar (1908), Estadística educativa de España (1974)
Agrictulture statistics 1865, 1891, 1917, 1962	Censo de la ganadería de España (1865), Avance de la riqueza pecuaria en 1891, Estudio de la ganadería en España (1917), Censo Agrario (1962) https://www.ige.eu/web/mostrar_paxina.jsp?paxina=002006&idioma=es
Rainfall data 1800-1930	Pauling et al. 2006 European Gridded Seasonal Precipitation Reconstructions https://www.ncdc.noaa.gov/paleo-search/study/6342

TABLE B2: Main contemporary sources

Description	Source
Demographic and educational data 1981, 1991, 2001, 2011	Spanish population censuses http://www.ine.es/dyngs/INEbase/en/categoria.htm?c=Estadistica_P&cid=1254734710990
Internal mobility 1998, 2003	Spanish population registry (Padrón) http://www.ine.es/dyngs/INEbase/en/categoria.htm?c=Estadistica_P&cid=1254734710990
Survey data on beliefs 2007-2018	Public Opinion and Fiscal Policy Survey, Centro de Estudios Sociológicos http://www.cis.es/cis/opencms/EN/2_bancodatos/catalogoencuestas.html
Register of schools and libraries 2018	http://abertos.xunta.gal/catalogo/ensino formacion/ /dataset/0257/centros educativos galicia https://rbgalicia.xunta.gal/es/rede/directorio mapa?field_biblios_prov_concello_tid=624
Register of local associations 2018	https://datos.gob.es/en/catalogo/a12002994 registro de asociaciones

Return to section 3 Data.

C Destinations and economic growth in Latin America

Galician migrants headed almost exclusively to three destinations: Rio de la Plata,[35] Cuba and Brazil. While these countries accounted for 97% of migrants in 1850-1930, their relative importance varied substantially from some areas to others and over time.

Province level data: The official migration statistics provide information on the destination of Galician migrants only for the 4 provinces and the periods 1885-95 and 1911-34.

#Name of sources #Overview of destinations and returns across provinces

Municipality level data: The only information of migrant destinations across municipalities comes from different historical sources compiled by Vázquez-González (1999). The sources include embarkation lists, desertion records, and notarial documents for the periods 1851-1860, 1901-1912 and 1920-1930. I aggregated the original data at the municipality and district level as displayed in Table C3. The data of 1851-1860 is not exploited given its limited coverage, nevertheless, the destination shares of 1851-1860 and 1900-1930 are highly correlated across districts.

---

[35]I use the term Rio de la Plata to refer to Argentina and Uruguay, as migrants could easily transit from one country to to the other.

TABLE C1: Official statistics of departures by province and country

| | 1911-1934 | | | | | 1911-1934 | | | |
	Total	Rio de la Plata	Cuba	Brazil	Other	% Rio de la Plata	% Cuba	% Brazil	% Other
Coruña	177,734	110,169	57,316	2,954	7,295	0.62	0.32	0.02	0.04
Lugo	141,213	75,798	61,966	1,273	2,176	0.54	0.44	0.01	0.02
Ourense	151,475	67,332	59,759	19,173	5,211	0.44	0.39	0.13	0.03
Pontevedra	151,586	100,714	22,648	23,409	4,815	0.66	0.15	0.15	0.03
Total departures	622,008	354,013	201,689	46,809	19,497	0.57	0.32	0.08	0.03
	1885-1895					1885-1895			
	Total	Rio de la Plata	Cuba	Brazil	Other	% Rio de la Plata	% Cuba	% Brazil	% Other
Coruña	69,853	26,168	35,846	5,717	2,122	0.37	0.51	0.08	0.03
Lugo	18,314	4,739	12,809	390	376	0.26	0.70	0.02	0.02
Ourense	21,235	3,446	6,999	9,877	913	0.16	0.33	0.47	0.04
Pontevedra	58,689	33,262	9,299	13,963	2,165	0.57	0.16	0.24	0.04
Total departures	168,091	67,615	64,953	29,947	5,576	0.40	0.39	0.18	0.03

Source: Return to page .

TABLE C2: Official statistics of returns by province and country

| | 1916-1934 | | | | | 1916-1934 | | | |
	Total	Rio de la Plata	Cuba	Brazil	Other	% Rio de la Plata	% Cuba	% Brazil	% Other
Coruña	106,589	37,799	54,884	1,522	12,384	0.35	0.51	0.01	0.12
Lugo	65,258	20,579	41,721	575	2,383	0.32	0.64	0.01	0.04
Ourense	75,107	22,388	42,470	4,796	5,453	0.30	0.57	0.06	0.07
Pontevedra	80,582	44,696	21,321	9,151	5,414	0.55	0.26	0.11	0.07
Total returns	327,536	125,462	160,396	16,044	25,634	0.38	0.49	0.05	0.08
	1887-1995					1887-1995			
	Total	Rio de la Plata	Cuba	Brazil	Other	% Rio de la Plata	% Cuba	% Brazil	% Other
Coruña	23,415	8,131	13,944	749	591	0.35	0.60	0.03	0.03
Lugo	5,299	725	4,353	68	153	0.14	0.82	0.01	0.03
Ourense	4,200	890	2,048	1,154	108	0.21	0.49	0.27	0.03
Pontevedra	20,429	12,482	3,683	3,733	531	0.61	0.18	0.18	0.03
Total returns	53,343	22,228	24,028	5,704	1,383	0.42	0.45	0.11	0.03

Source: Return to page .

TABLE C3: Main migrant destinations across Galician districts, 1900-1930

District name	Total	Rio de la Plata	Cuba	Brazil	% Rio de la Plata	% Cuba	% Brazil
Allariz	130	51	28	51	0.26	0.37	0.37
Arzua	581	478	100	3	0.18	0.81	0.00
Bande	105	62	26	17	0.25	0.59	0.16
Becerrea	346	269	76	1	0.22	0.78	0.00
Betanzos	711	571	140	0	0.24	0.76	0.00
Caldas	429	410	14	5	0.04	0.96	0.00
Cambados	887	848	31	8	0.03	0.96	0.00
Canhiza	114	95	6	13	0.08	0.79	0.13
Carballinho	193	130	46	17	0.34	0.58	0.08
Carballo	749	654	95	0	0.12	0.88	0.00
Celanova	105	62	37	6	0.44	0.51	0.05
Chantada	810	532	273	5	0.34	0.65	0.01
Corcubion	731	612	117	2	0.16	0.84	0.00
Corunha	1,059	801	258	0	0.24	0.76	0.00
Estrada	175	145	19	11	0.08	0.90	0.02
Ferrol	432	261	155	16	0.37	0.59	0.04
Fonsagrada	391	281	108	2	0.28	0.72	0.01
Xinzo	142	40	32	70	0.33	0.24	0.42
Lalin	737	533	190	14	0.25	0.73	0.02
Lugo	823	503	312	8	0.38	0.61	0.01
Mondonhedo	528	378	147	3	0.28	0.72	0.01
Monforte	231	88	142	1	0.61	0.38	0.00
Muros	184	146	36	2	0.21	0.78	0.01
Negreira	260	165	95	0	0.33	0.67	0.00
Noia	369	350	18	1	0.18	0.82	0.00
Ordes	341	231	108	2	0.32	0.68	0.01
Ourense	335	207	103	25	0.37	0.56	0.07
Ortiguera	235	66	169	0	0.72	0.28	0.00
Padron	545	503	34	8	0.05	0.95	0.01
Pontevedra	245	205	34	6	0.21	0.77	0.01
Pobra de Trives	149	63	61	25	0.51	0.35	0.14
Ponte Caldelas	145	101	4	40	0.04	0.70	0.26
Ponteareas	161	148	9	4	0.10	0.86	0.05
Pontedeume	449	240	204	5	0.51	0.48	0.01
Quiroga	132	56	73	3	0.56	0.42	0.02
Redondela	102	84	9	9	0.09	0.76	0.15
Ribadavia	130	96	30	4	0.58	0.40	0.02
Ribadeo	271	167	102	2	0.37	0.62	0.01
Santiago	571	444	36	91	0.07	0.78	0.15
Sarria	415	253	158	4	0.38	0.61	0.01
Tui	288	248	13	27	0.07	0.83	0.10
Valdeorras	253	93	134	26	0.54	0.36	0.10
Verin	455	128	47	280	0.14	0.27	0.59
Viana do Bolo	144	81	31	32	0.30	0.50	0.20
Vigo	359	308	30	21	0.15	0.81	0.04
Vilalba	484	269	207	8	0.43	0.55	0.02
Viveiro	453	303	146	4	0.33	0.66	0.01
Total	17,884	12,759	4,243	882	0.24	0.71	0.05

Economic growth at destination: The Maddison Project Database 2018 provides data of GDP per capita in the main Latin American countries over the period 1850-1930. To construct a measure of economic growth at the average destination I weight each country with the municipality level data from Vázquez-González (1999). I rely on the 1900-1930 data because it is the most exhaustive and highly correlated with that of previous periods. I construct alternative measures with data aggregated at the distrcit level (1900-1930) and the province level data (1885-1895) from the official migration statistics.

FIGURE C1: Economic growth at the average migrant destination, 1877-1930
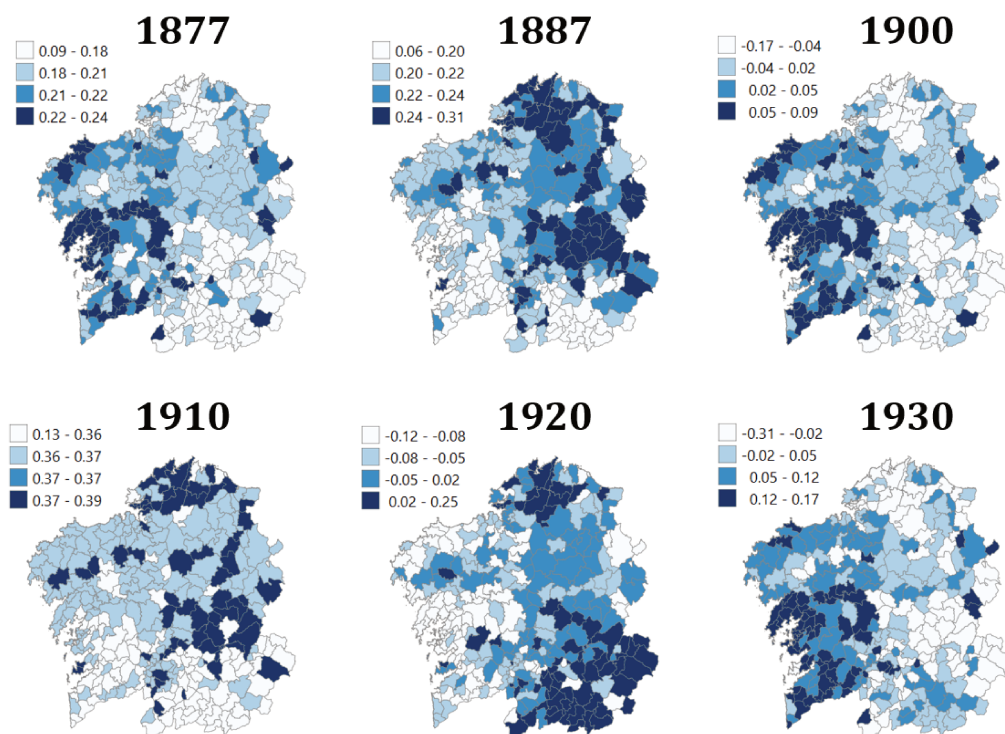
Note: the figure displays GDP per capita growth at the average migrant destination by decade using as weights the main destinations in 1900-1930. For example, if 80% of migrants from a municipality headed to Rio de la Plata and 20% to Cuba, the GDP per capita growth of each country is weighted accordingly. Return to page 8 or 14.

D Weather shocks at origin

In 1860-1930, over 80% of Galicians worked in subsistence agriculture. As a result, weather conditions could have a large impact on households' earnings as well as on migration decisions. I rely on historical rainfall data to construct measures of local economic shocks in this period.[36]

Weather data: I use data from the European Gridded Seasonal Precipitation Reconstructions (Pauling et al. 2006) which contains yearly information of precipitation at a high-degree resolution from 1500 on wards. Its fine resolution (0.5° x 0.5°) yields a grid of 24 cells in Galicia with an average approximate size of 50km^2 (53km x 43km).

Rainfall shocks: I construct a measure of rainfall variability in the harvest season by computing the standard deviation of precipitation in each grid cell and decade:

$$SDR_{gt} = \sqrt{\frac{1}{10} \sum_{i=t-9}^{t} (r_{gi} - \bar{r}_{gt})^2}$$

where r_{gi} represents total precipitation in cell g and year i during the Fall and $\bar{r}_{gt}$ the average total precipitation in cell g during the decade preceding t ($\bar{r}_{gt} = \frac{1}{10} \sum_{i=t-9}^{t} r_{gi}$).

Note that since I will use this measure in a panel setting, the municipality fixed effects will capture the fact that some areas have higher variability than others on average. SDR_{gt} should then be interpreted as changes with respect to the average variability of a location.

In the cross-sectional analysis I compute the standard deviation difference with respect to the mean in each grid cell and decade as follows:

$$ZR_{gt} = \frac{1}{10} \sum_{i=t-9}^{t} \frac{|r_{gi} - \bar{r}_{gt}|}{SDR_{gt}} \ .$$

---

[36]While both temperature and rainfall are important determinants for agricultural yields, there is much more variation in rainfall than temperatures across the different areas of Galicia.

FIGURE D1: Rainfall variability in the previous decade, 1887-1930
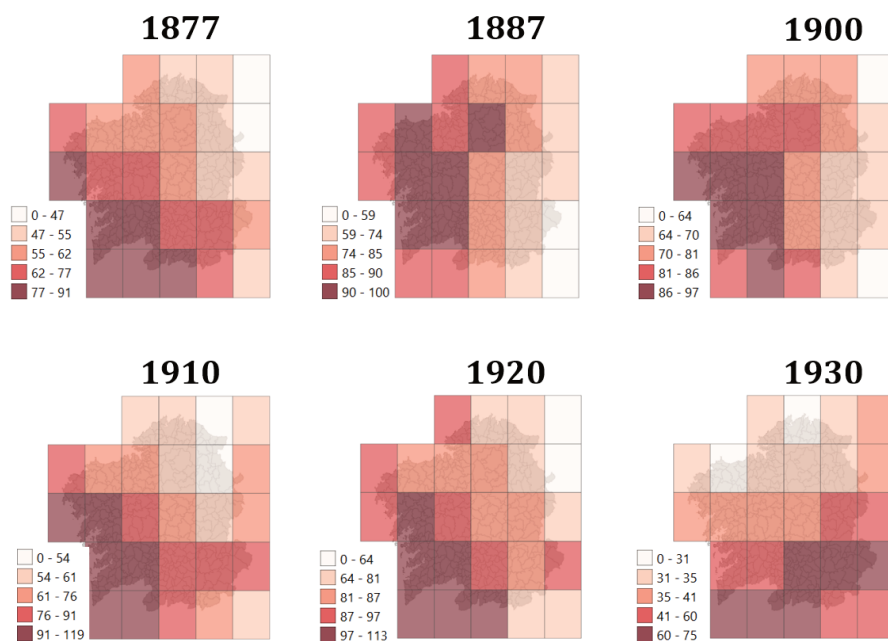
Note: the figure shows the variability in rainfall in every decade (standard deviation). Darker colours represent higher variability.

FIGURE D2: Rainfall variability in the previous decade (relative), 1887-1930
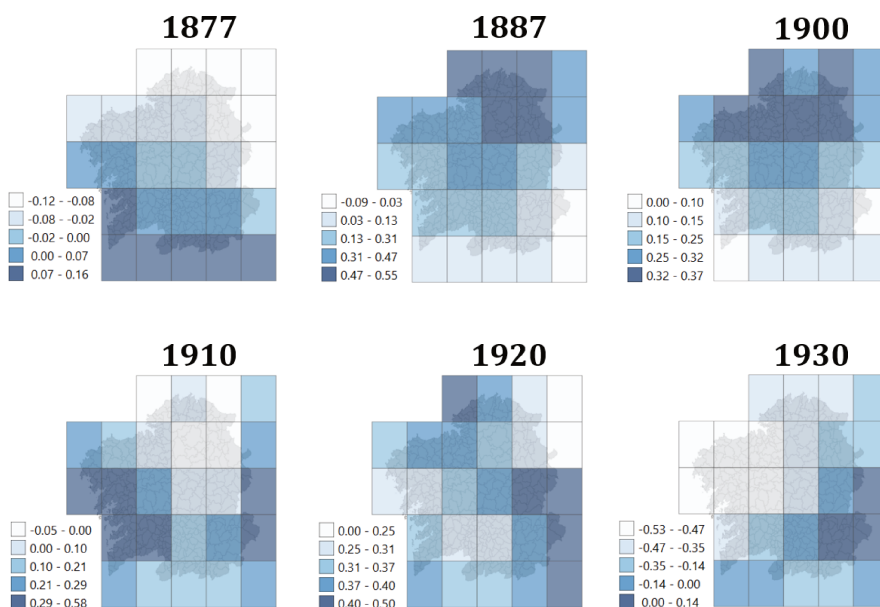
Note: the figure shows the variability in rainfall in every decade relative to the standard in each location (z-score), using as reference the average precipitation during 1860-1930. Darker colours represent higher variability.

Alternative data & validation: Since maize is harvested at the end of September/ beginning of October my weather shocks measure relies on rainfall variability in the Fall. Yet, a priory it is uncertain which periods or shock measures could be more relevant to explain migration. To give additional support to this choice I rely on data from the last release of CRU TS (CRU TS v. 4.03) which contains monthly information of rainfall from 1901 onward at the same resolution (0.5° x °). This monthly data allows me to compute multiple measures of rainfall shocks (e.g., variability/excessive/insufficient rainfall in each month, group of months, etc.). I therefore constructed 240 different variables of rainfall shocks and implemented a Lasso procedure particularly adapted for panel settings, the Cluster-Lasso (Belloni et al., 2016). Using this procedure I found that the variables that better explain changes in the share of missing men in 1910-1930 were rainfall variability in October and excessive precipitation in October (e.g., number of months in the decade with rainfall higher than one standard deviation above the mean).

E   Distance to the port and migrant networks

Background: The first migration wave in the 1850s affected mostly areas closer to the transatlantic ports. Local agents working for shipping lines took an active role spreading information and facilitating migration in rural municipalities with no prior experience, expanding their catchment areas inland over time.[37] However, the diffusion of information was hampered by the poor transportation infrastructure that had historically isolated municipalities from each other. In the second half of the nineteenth century, the distance to a transatlantic port was a major determinant of migrant flows, not only due to the information channel but also to the high transportation costs implied by individuals having to walk until the ports.

Computing the distance to the closest port: The most direct way to measure the distance to a port is to estimate the linear shortest path.[38] Yet, in a context with numerous natural barriers (e.g., mountains, rivers, etc.) as in Galicia, there can be an important mist-match between "as the crow flies" distances and distances following actual roads. To account for this I manually draw and geo-referenced the network of the primary walking paths in the 1880 using historical maps of Galicia. Using this network I can estimate the walking distance in km to the closest transatlantic port. To test the validity of this measure, I also rely on the shortest walking path to these ports today based on Google maps.

---

[37]These agents were important to familiarize individuals with the opportunities abroad, provide them with the necessary documents (e.g., passports, tickets, etc.) and even finance the trip.

[38]One example is Karadja & Prawitz (2019) where they use the great-circle haversine formula to take into account the shape of the Earth.

FIGURE E1: Transatlantic ports and primary roads in Galicia in 1884
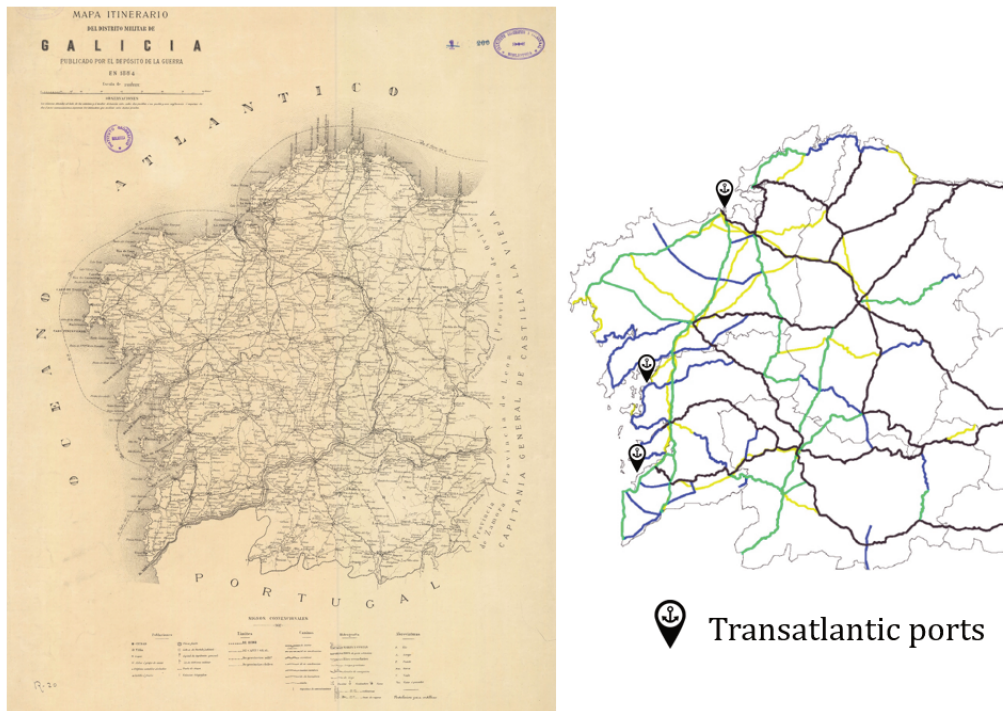
Note: The symbol of an anchor indicates the ports shipping boats to the Americas. From North to South: Coruña, Carril/Vilagarcía and Vigo.

FIGURE E2: Comparison between linear and walking distance
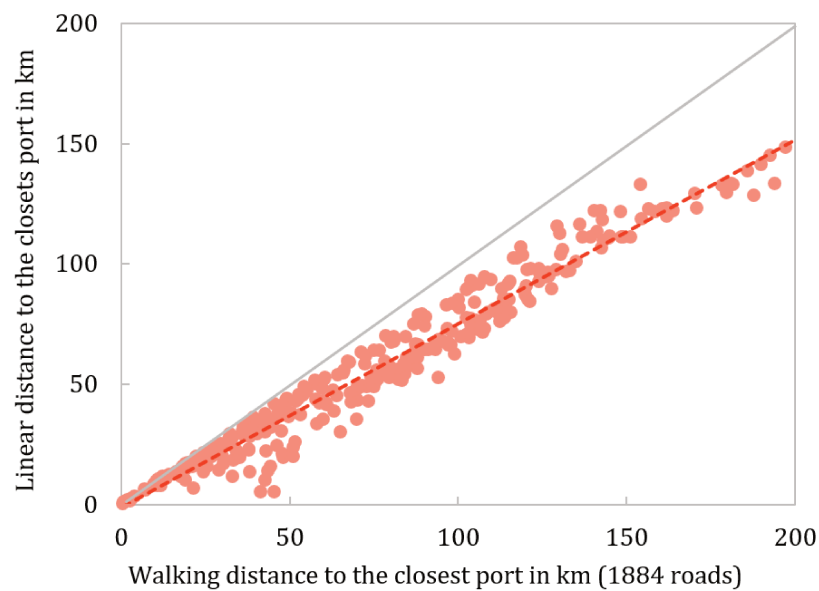
Note: The figure shows the linear distance (in km) to the closets transatlantic port and the actual walking distance (in km) following the road network in 1884.

F   Spatially correlated errors (Kelly, 2019)

In a recent working paper, Kelly (2019) highlights that spatially correlated errors can lead to distorted significance levels in spatial regressions. In particular, he shows that randomly generated noise variables with a certain degree of spatial correlation can be significant predictors of historical and contemporary (spatially correlated) variables. Revisiting 27 papers published in top journals, he concludes that: *"only about one quarter of the persistence results that we examine are robust after we take account of the possibility that their regressions might be fitting spatial noise."*

The paper proposes a simple procedure to investigate the role played by spatial noise in influencing the significance of regression results. In a nutshell, it consists in generating random noise variables with spatial correlation and using them as explanatory or dependent variable in the analysis. Figure F1 is an example of a spatially correlated random noise variable in a square grid of 35x35 cells.[39] I generate one thousand grids like Figure F1 and mapped them to my database of Galician municipalities to perform the tests suggested.

FIGURE F1: Randon noise variable with spatial correlation
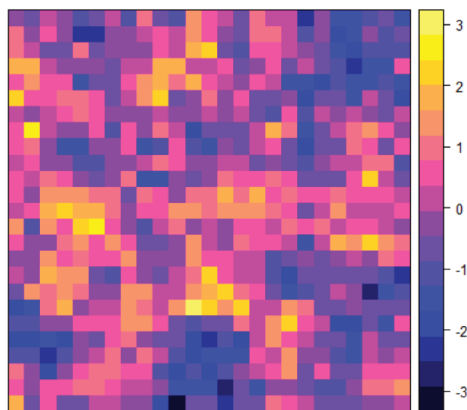
Note: The figure shows a random variable in a 35x35 grid, with a spatial correlation of range 2.

I first regress the share of individuals with secondary education in 1981 (the dependent variable in Figure 13 and Table 5) sequentially on these thousand random noise variables with spatial correlation. I obtain a significant coefficient at the 5%, 1% and 0.1% level in 12.9% of the cases, 4.6% and 0.7% respectively. These point estimates are extremely small and heavily centered around zero (95% of them range between -0.004 and 0.004) while my ('true') estimated effect is 0.14 (Table 5) .

---

[39]I use these dimensions so that they correspond approximately to the size of municipalities. I also try generating alternative maps with different ranges of spatial correlation and obtaining similar results.

Secondly, I replicate equation (5) a thousand times using as dependent variable with a randomly generated spatially correlated noise variable and instrumenting the share of missing men (1900-1930) with the adult sex ratio in 1860. The results are displayed in Figure F2 in comparison with those obtained by Kelly (2019) for other studies. Overall, despite a substantial spatial correlation in the explanatory variable and instrument, the robustness of my analysis fares relatively well compared to previous work. It is important to stress once again that this comparison only refers to the degree of potential distortion in significance levels, and not to the size of the estimates effects, that in my case are very large. The point estimates in Table 5 range from 0.14 to 0.46. Having noise as a dependent variable and performing my instrumental variables strategy one could obtain a point estimate higher than 0.14 in less than 17.5% of the cases and as large as 0.46 in less than 2.3% of the cases.

FIGURE F2: Ability of persistent regressions to explain random noise
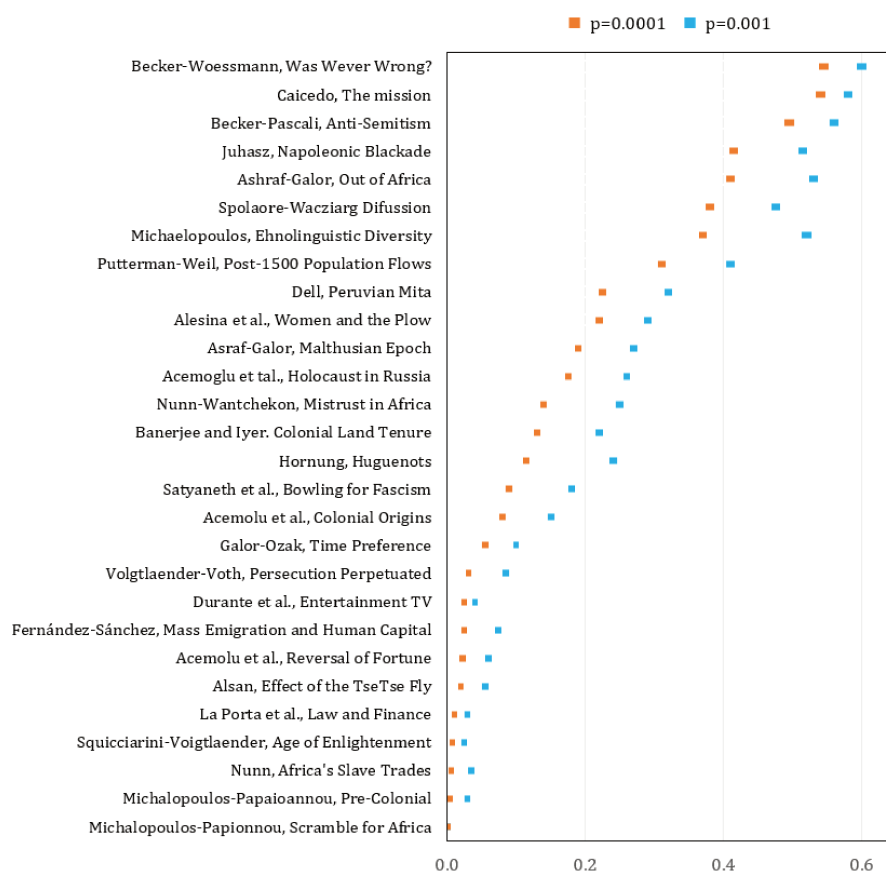
Note: The figure replicates Figure 11 in Kelly (2019) adding the results for this paper. It represents the fraction of simulations where the persistence variable explained spatial noise at a significance level of 0.001 or 0.0001.

G   Anecdotal evidence

This section gathers direct anecdotal evidence from Galician migrants about the importance of education, migrants' associations and schools, and other aspects. The original quotes in Spanish/Galician can be found here.

Migrants realization of the importance of education [Return to page 27].

1. "Each of them (migrants) recalls the poor education received and stands up, proclaiming that they want that future generations are nurtured with the power of the most useful and needed knowledge to go around the world." - El emigrado, A Estrada, no. 7-7-1923

2. "It is crucial that we realize the deficient education we received, and acknowledging how difficult it is out there without it, that we transmit it to our brother and sisters so that can enjoy better lives."- «El Progreso de Coles». Pro galicia. Vol. I, n. 2, 1912 (p.19)

3. " [...] understanding the need to educate people at origin [...] not only to facilitate basic knowledge to those migrating, but to diffuse the power of education back home, the only way to free ourselves from exploitation and elites' abuses [...]" - Álvarez Gallego (1906, p.313)

4. "[...] In Cuba they never asked me if I knew how to pray, they always asked if I knew how read and write and since I said yes, I got better and better jobs." - Pereiro (2012:3)

5. "Schools in Galicia are the result of love and patriotism from good heart people who know the power of education to conquer good, wealth and freedom. And that is what (migrants) want for their brothers and sisters, [...] they dream that in the land they yearn for, a more educated society will lead to political and economic progress." - Nova Galicia, n. 472 de 1915

6. "[...] With what will be taught there (migrants' schools) people will have enough to fare better than those migrating in the past without knowing how to read, write or count."- «Mi opinión». El Faro de Veiga, Colegio San Adrián, 1914 (pp.38-39)

7. "Schools are the best investment, with more immediate and future outcome, that we migrants can do for the economic revival of Galicia - [...]" - Alonso Rios, 1929

8. "[...] We Galician migrants, thanks to the educational associations we are supporting with so much effort, we aim to bring to our hometowns the highest education possible, educating our children and building schools not only to prepare future generations for migration, but with the aim to take advantage of all the wealth our region has. [...]" - José Loureiro García: «Siempre adelante», Galicia. Revista Regional Ilustrada. Órgano de la Colonia Gallega y Sociedades Regionales de Cuba, A Habana, no. 40, 30 de setembro de 1922, s.p.

   Explicitly transmitting higher value for education.

9. " [...] The driving force of social transformation in our country will be knowledge and the education of our people through newspapers, magazines and books, conferences and meetings, but most importantly, schools." - Constantino Horta: « Cultura y regionalismo II», Ecos del Eume, Pontedeume, no. 181, 26 de maio de 1912

10. "It has been two years since the foundation of this association [...] aiming at contructing and financing [...] schools with the best conditions and all the pedagogical innovations needed [...] If we want a free country and responsible citizens, we have to start by disseminating education among the young generations and making our brothers and sisters understand their rights [...]"-Junta directiva Sociedad de Instrucción Alianza de Vilamarín (1912, Habana)

11. "–¿Emigration? - asks a friend. But, you think emigration is evil? All the money Galicians make in the Americas comes back here, fueling our industry. And it is not just money what migrants make circulate between us, it is the spirit of progress and tolerance [...], Galician migrants are modernizing Galicia. Do not be a fool. Emigration is good." - Julio Camba: «La emigración es un bien», Eco de Galicia, A Habana, no. 115, 21 de marzo de 1920, (p.8)

12. "In Cuba my father worked in the rail company because he knew how to read and write, and he always said that made him improve [...] That is why he was obsessed that we could not miss a single day of school. [...]" - Extraído de la vida de Eurisna Campos Cudeiro (Amoeiro, OU).

13. "In their letters, often written by others' hands, what do they (migrants) tell you? [...] Do not they advise you, with their sad regret, that you should go to school? So yes, my beloved neighbors, I encourage you, leaving the laws aside, that you should send your kids to school as much as possible."- «Un bando que honra a un pueblo y a su alcalde», Galicia. Revista del Centro Gallego, Bos Aires, no. 216, decembro de 1930, (p.6)

14. "[...] Greetings to my beloved brothers and sisters, I hope the work is not to harsh, and to Pepe and all the others at home, make sure Aquilino studies well." - Cayetano Borrajo (La Habana, 21-7-2018)

15. "I know sometimes you have no option but missing class because you need to help your mothers at home, but do please ask them to let you come to school, to not miss class; tell them that we do not live only on present and palpable things." - Discurso de clausura do curso académico pronunciado pola mestra da escola de nenas de Ares, creada pola Sociedad Instructiva Redes-Caamouco de La Habana - Memoria de la Sociedad Instructiva Redes-Caamouco, 1922

16. "I would like to ask public authorities to enforce the educational law that obliges children to attend school, and that kids are examined by doctors to prevent the spread of contagious diseases ." - Conclusión do Primeiro Congreso da Federación de Sociedades Gallegas Agrarias y Culturales, Buenos Aires, El Tea, Puenteareas, 1922.

17. "CHAPTER ONE NAME AND AIM OF THE ASSOCIATION [...] Article 5 We urge the corresponding public authorities to make education compulsory between the ages 7 to 14, setting penalties for parents who do not comply to these rules." - Estatutos de la Sociedad Pro-Escuelas en Paradela y sus Contornos. Ayuntamiento de Meis en Pontevedra, Buenos Aires, Imp. F. Vilanova, 1912, (pp. 3-4)

18. "WE RECOMMEND to all families that, [...] either individually or collectively, try by all means to avoid the payments to the church and elites and in turn use that money to purchase materials for schools, to pay teachers and everything that could improve education [...]" - Teo. Órgano Oficial de la Federación Residentes de Teo en Sud América, Bos Aires, no. 41, 15 de setembro de 1913, (p.1)

Praising the role of migrants' associations.

19. "From all the exceptional actions that the Galicians who are missing did, the greatest without doubt, has been the creation of associations that gathered funds for the construction of schools, and for the diffusion of culture in their home country." - «Una gran obra de cultura para Galicia», Vida Gallega, Vigo, 1912.

20. "I remember that schools, for long time, were critized as some said they encouraged emigration. They said they were centers to train future migrants. Yet, the reality was that Galicia lived in misery and backwardness. The first and immediate objective of a migrant was to contribute to improving the village or parish where they grew up, among other things, because that village or parish was the only place of their own, the only part of Galicia they knew." - M. Meilán en: Samuelle, C.,80:1993

21. "This institutions (migrants' associations) will rise the culture of Galicia to an unprecedented height, and improve the doncidionts of our migrants, facilitating their labor market insertion in the Americas, and educating the people in our fields, rising their moral and public wealth." - Alianza Aresana de Instrucción, Vida Gallega, 1911.

22. "The wave continues, the enthusiasm is everywhere, educational migrant associations are spreading so fast that very soon will be no Galician municipality without a school constructed and financed by our brothers emigrated to the Americas." - Sueva, Habana, 1910.

23. "Galicia, for its peculiar geography, is in real need of diffusing basic education. We need to found many schools, loads of them -good and less good-, to bring education to all little towns in the mountains and the valleys [...] War against illiteracy, that's what is most pressing thing; Galicia without a single illiterate, that should be the aim of all our aspirations." - Nicasio Pajares: « El magno problema», Fomento de la Instrucción Gallega. Órgano de la Sociedad Pro escuela en Bandeira, no. 4, 15 de novembro de 1909, (pp. 2-3)

24. "[...] I will never have enough words to celebrate the workers from Ares, my beloved colleagues, whom with their modest earnings, barely enough to sustain themselves in cities as expensive as Havane, they have managed to found and finance a school for boys and soon one for girls in their hometowns, equipping them with all materials, and with all the hygiene conditions and modern necessities [...]" - Manuel Curros Enríquez: «Bien por los aresanos!», Galicia, A Habana, no. 30, 25 de xullo de 1905, (p. 8)

Migrants hard work and high savings.

25. "Most of the people were foreigners who came to earn some money and return. Me too, I thought I would spend here more than three or four years, I thought I would come back soon [...] But while I was in the Americas, everything was saving. It was a really different life."-Buenos Aires, febrero 1994. Interviews from Samuelle-Lamela (2000)

26. "[...] Galicians migrated to earn some money and had no option but saving a lot. Generally they could come here to marry spend three or four years and come back. You could identify Galicians by their skin color, so pale because of spending all their time working behind the counter." - Ricardo Flores, Buenos Aires 1994. Interviews from Samuelle-Lamela (2000)

27. "I went to work in a restaurant, from 5 in the morning to 9 at night, only resting one day pwer week and one hour for lunch and another for dinner." - Montevideo, agosto 1988. Interviews from Samuelle-Lamela (2000)

About conditions in migrants' schools.

28. "This is the sshool "Unión Hispano Americana Valle Miñor´´. We have more 300 kids from 7 to 15 years old. Education is completely free; and the poor, who are more than 160, are offered free books, clothes [...] During the most favorable season, we offer night lectures, having 60 young adults attending. These classes are also free of charge." - «Escuela Americana», Boletín de la Unión Hispano Americana Pro-Valle Miñor, Bos Aires, no. 40, xaneiro de 1913, (p.1)

29. "[...] Education will be free o charge for everybody who needs it, we will give them books, ink, paper, and all necessary materials. To enroll in this school the only conditions is to be 6 years or older and [...] All students that without justified reason (such as sickness, bad weather, impassable roads) do not attend for a week will have their parents notified and will be expelled from the school. Everybody who is not from Mera is allowed to come here, send their kids, brothers or anyone in need of education." - Reglamento para el interior del Colegio de la Sociedad de Instrucción PUENTE DE MERA

30. "The basics (of teaching practices), according to the aspirations of the founders of or association "Ferrol y su Comarcar", will be to use the most intuitive and practical methods possible to transmit the basic knowledge in modern life [...]" - Sociedad Ferrol y su Comarca (1911, p.5-6)

31. "Making education rational, practical, harmonic and progressive, training our students to think, reflect, and to do research on their own [...] Considering the skills and conditions of every kid, making our lessons clear and precise, using all kinds of intuitive methods, relying on materials in our class, [...], so that education is entertaining, interesting, varied

and likeable, making kids put to test the knowledge they have acquired [...]" - Boletín de la Unión Hispano–Americana pro Valle Miñor (1914)

Criticism of migrants' schools because they might foster more migration.

32. "It helpful and interesting the endeavor of Galician migrants in America financing schools in Galicia. But, would it not be more fruitful and a better example if they led the State alone to build the schools, being its responsibility? [...] There are also some of those schools that, with good intentions, end up being dangerous and lethal, since they are centers where Galician identity is lost and were they instill on children the desire to emigrate, instead of transmitting them love for what is ours." - Antón Villar Ponte: «Cartas da vila. A civilización», A Nosa Terra, A Coruña, no. 96, 5 de agosto de 1919, (p.5)

33. "Meanwhile in our country, nothing is done for the "Galician" school, migrants, with a wrong approach but good intentions, waste our money in financing schools that have a bad orientation for Galicia. Those schools supported with money from the Americas do not only make children lose their Galician identity, they even guide students towards migrating, they even adopt history and geography books from these countries." - «A creación da escola galega, como toda laboura galeguizante debe contar conorganizacións que cordinen e metodicen a acción de todol-os galeguistas», A Fouce, Buenos Aires, 1931

34. "[...] Galician in the Americas finance schools, that is true. There, Galician kids learn what is good and evil, what they should like and dislike, they learn geography, language, history, all of that without the spirit of Galicia. The schools financed by migrants have the same orientation as the schools financed by the State and thus collaborate to the loss of Galician identity [...]" - Xan de Pazos: «Algunhas refreusiós ós galegos da América», A Fouce. Periódico Galego, Bos Aires, no. 3, xuño de 1926, (p.1)

35. "The problem is that too much money is invested in the construction of these schools, to the detriment of the main aim which is teaching and education [...]" - M. Nodar Magán: «Un consejo», El Emigrado, A Estrada, no. 22, 16 de xuño de 1921, (p.1)

Chained migration and importance of contacts.

36. "Those standing out were the ones emigrating to Cuba at the end of the nineteenth century and coming back looking wealthy. That instilled in others the desire to get rich themselves.

Above all, they were different for their appearance: they came back with a car, dressed like the elite and wearing many golden things."

37. "I had three seasons that motivated me to leave. We are a numerous family struggling financially because only my father had a formal salary, while the rest of us worked in the field [...] Something else that made me dream was when I saw other migrants coming back from the Americas with modern cars, and they arrived and refurbished their houses, and people said "Look the *Brazilian* has arrived, full of money!" And I was young and with that hope though "What am I doing here?" [...] And thirdly I already had there two older brothers and one uncle. They had already settled and had money. They sent me an invitation letter and I went as a mechanic [...]"-Luis Vilariño Garrido (Brazil)

38. "In my time it was very simple. Almost always there was a pionneer, he would get a job, settled, and just after, he would begin calling others back home: brothers, friends, cousins, neighbors [...] relationships between people from the same place were very important, because besides helping each other, it was the way to access the labor market." - Manuel Meilán, interviewed by Samuelle-Lamela (2000)

39. "Almost everybody had some relative here (in Latin America), and in that moment of high migration there was plenty of work here, and I think they believed that relatives from Spain would work harder that having employees from there." - Montevideo, octubre 1986. Interviews from Samuelle-Lamela (2000)

40. "Far from home, the first thing they did was to look for a neighbor, recreating their parish of origin there, and that's how they founded local associations. Those who succeded came back to Galicia and brought other neighbors to work in some business (in Latin America). That is how there were bakers from an area, builders from another, taxi drivers from Chantada, barman from Ourense, choffers from Bergantiños, butchers from Val Miñor [...]" Montevideo, 10 de octubre 1986. Interviews from Samuelle-Lamela (2000)

41. "My grandfather and uncle had come earlier and they always told me: from Argentine to paradise."- Buenos Aires, marzo 1994. Interviews from Samuelle-Lamela (2000)

Other.

42. "The only trade between Galicia and the colonies is precisely migration to the latter [...]"- Spain general correspondence, FO 72/702, 25/3/1946. [cited by Vázquez-González (1999)]

43. "Excluding a minority that comes back to Spain within a year, most of them stay two or up to five years and then come back [...] The idea that there is short-lived circular migration to the republics of Plata is simply ridiculous."- Consejo Superior de la Emigración, 1916. [cited by Vázquez-González (1999)]

44. "From Galicia only men emigrate. Galician women very rarely come to Portugal, they stay home raising and educating their children."- As Farpas, Ortigâo, R. 1950. [cited by Vázquez-González (1999, pp.107)]

H   Brief description of the Galicia economy, 1800-1930

During the XIX century and until 1920, Galicia was the poorest region of Spain. In this period, income per capita levels were around half of the country's average (Rosés, Martínez-Galarraga and Tirado, 2010), the vast majority of the population worked in the primary sector (close to 90%) and urbanization levels were very low (less than 10% of municipalities had a population above 10 thousand inhabitants).

As described by Alonso-Álvarez (2010), most land was owned by landlords (´´fidalgos") but exploited by peasants under long-term concessions ("contratos forales"). While these agreements made peasants *de facto* quasi-proprietaries, they were obliged to pay a yearly in-kind contribution ("foros") on top of other taxes to the Church and local elites. This property system influenced the type of plots (small and geographically dispersed) and farming practices (extensive use of the plough and little investments in machinery or fertilizers). The main crops were maize, wheat and rye, whose production was devoted to household consumption and the local market. Livestock was also very common, contributing to producing natural fertilizer and being a crucial link in agriculture production. Communal lands, such as forests and meadows, were exploited for grazing, to obtain wood and other inputs (e.g, gorse for compost). Overall, Galician agriculture was very inefficient and harvests barely supported subsistence level.

The low agricultural productivity forced households to engage in supplementary activities such as rural domestic industries (e.g., fish salting, mule driving, blacksmithing, weaving and tanning) and seasonal migration to the harvest in Castilla and Portugal (Farías, 2010; Palmás, 1976). These activities did not require specialization and could be carried out during periods of less intense agricultural work. The output produced by rural industries could be sold in other parts of Spain through the seasonal migrations (Alonso-Álvarez, 2010).

A central aspect of the Galician economy was the high degree of population dispersion and the poor transportation infrastructure. Population was scattered into several settlements ("parroquias") that constituted a local market, with an average of twelve settlements per municipality. As a result, peasants did not need good roads for their daily activities but relied instead on a comprehensive network of ground walking paths (Barreiro Gil, 2001). Yet, the deficient transportation infrastructure became an obstacle for establishing an integrated regional market and contributed to the economic backwardness of Galicia (Correa, 2008; Barreiro Gil, 2001). The bad state of roads, especially after adverse rainfall events, increased transportation costs making it unprofitable to

sell products far away (Correa, 2008) and leading to a system in which rural markets operated in autarchy. The exceptionally high number of fairs in Galicia (more than 4000), rather than economic dynamism, exemplifies the atomization of the market (Barreiro Gil, 1983; Correa, 2008). The maintenance of paths depended on municipalities, but since neighbors already supported a high tax burden, it was very difficult to increase taxes (Barreiro Gil, 2001). Note that although dismantled by the nineteen century, institutional barriers such as tolls in crossing points had also contributed to the isolation of municipalities and of Galician from the rest of the country.

Thorough the nineteenth century, the Galician economy evolved very slowly. While the State progressively replaced the Church and landlords as main tax collector, fiscal pressure did not decrease. The modest textile industry that had expanded during the first decades of the century went bankrupt during the 1830s. Other agricultural crises would occur in the following decades. The canning industry emerged in the mid-century, driving other related sectors (e.g., fishing, oil, tin, and machinery) but was marginal in terms of employment (Alonso-Álvarez, 2010). Cattle was the main and almost sole Galician export, raising rapidly in 1860s but collapsing after the closure of the British market at the end of the nineteenth century (Barreiro Gil, 1983). Most railroads opened during the 1870s and 1880s, connecting Galicia with the rest of Spain. The access to railroads was crucial helping redirecting the fishing and cattle sectors towards the domestic market during the twentieth century (Barreiro Gil, 1983). It also contributed to a more integrated regional market and to the reduction of price disparities.

In the 1900-1930 period, Galician agriculture experience several changes that made it more productivity. First, most farmers became full owners, being thus redeemed from paying rents (Alonso-Álvarez, 2010). Getting access to the land was crucial for the adoption of technological innovations such as better seeds, chemical fertilizers, machinery, etc. This innovations spread very rapidly thanks to agrarian unions and the access to credit (Fernández-Prieto, 1997). For instance, by the 1930s the thrasher had completely replace the traditional threshing mallets. One of the greatest changes occurred in the sector of ranching thanks to the introduction of hygiene measures, disease control and new breads of cattle (Alonso-Álvarez, 2010). The agricultural sector became more integrated in the Spanish market with cattle and wood being the main products. New industries related to fishing activities expanded during the 1920s, including shipyards, sawmills, metallurgy, etc. These economic changes fostered an incipient urbanization and encouraged people to move from rural areas to the emerging cities. Nonetheless, Galicia remained being one the Spanish regions more underdeveloped with income per capita levels of 60% of the average in 1930.

Appendix References

ALONSO ÁLVAREZ, L. 2010. La economía de galicia, una panorámica, c. 1750-2010. *Historia contemporánea* .

BANDIERA, O., RASUL, I., AND VIARENGO, M. 2013. The making of modern america: Migratory flows in the age of mass migration. *Journal of Development Economics* 102:23–47.

BARREIRO GIL, M. J. 1983. La generalización de la produccion de mercancias y la modernizacion productiva de la agricultura en galicia, 1876–1976. *Revista de Historia Economica-Journal of Iberian and Latin American Economic History* 1:133–146.

BARREIRO GIL, M. J. 2001. El atraso económico y el sistema de transporte terrestre en la galicia del siglo xix. *Revista Galega de Economía* 10:0.

CORREA, H. S. 2008. Obstáculos estructurales al desarrollo de los intercambios comerciales en la galicia de la edad moderna. *Ohm: Obradoiro de Historia Moderna* .

COSTA RICO, A. 1989. Escolas e mestres. a educación en galicia: da restauración á segunda república. *Santiago de Compostela: Xunta de Galicia* .

DIEZ-MINGUELA, A., MARTINEZ-GALARRAGA, J., AND TIRADO-FABREGAT, D. A. 2018. Regional income inequality in spain 1860–2015, pp. 81–103. *In* Regional Inequality in Spain. Springer.

FERNÁNDEZ-PRIETO, L. 1997. Selección de innovaciones en una agricultura atlántica de pequeñas explotaciones. galicia 1900-1936. la adopción de las trilladoras mecánicas.

KELLY, M. 2019. The standard errors of persistence.

LLORDÉN MIÑAMBRES, M. 1998. Las asociaciones de inmigrantes españoles en américa. algunas respuestas a los desequilibrios y carencias de la emigración a ultramar. *Exils et migrations ibériques au XXe siècle* 2:79–130.

NARANJO OROVIO, C. 1988. Del campo a la bodega: Recuerdos de gallegos en cuba (siglo xx). *Sada, Ediciós do Castro* .

NÚÑEZ SEIXAS, X. M. 1998. Retornados e inadaptados: el" americano" gallego, entre mito y realidad (1880-1930). *Revista de Indias* 58:555–593.

Palmás, R. 1976. La emigración. *FABRA, Gustavo, et al. Los gallegos. Madrid, Istmo* .

Pauling, A., Luterbacher, J., Casty, C., and Wanner, H. 2006. Five hundred years of gridded high-resolution precipitation reconstructions over europe and the connection to large-scale circulation. *Climate dynamics* 26:387–405.

Peña Saavedra, V. 1991. Éxodo, organización comunitaria e intervención escolar: la impronta educativa de la emigración transoceánica en Galicia. Xunta de Galicia, Secretaría Xeral para as Relacións coas Comunidades Galegas.

Rico, A. C. 1989. Escolas e mestres: a educación en Galicia: da Restauración á IIa República. Servicio Central de Publicacións, Consellería da Presidencia e . . . .

Rosés, J. R., Martínez-Galarraga, J., and Tirado, D. A. 2010. The upswing of regional income inequality in spain (1860–1930). *Explorations in Economic History* 47:244–257.

Samuelle Lamela, C. 2000. La emigración gallega al Río de la Plata. PhD thesis, Universidad Complutense de Madrid.

Sánchez Alonso, B. 1995. Las causas de la emigración española, 1880-1930. Alianza Editorial Sa.

Vázquez González, A. 1999. La emigración gallega a américa, 1830-1930. *Memoria de doctorado inédita, Facultade de Ciencias Económicas e Empresariais, Universidade de Santiago de Compostela, Santiago de Compostela* 2.

Chapter 2

Més Que Un Club: Football, National Identity and Political Preferences

Més Que Un Club: Football, National Identity, and Political Preferences[*]

Abstract

Few things are as powerful as sports when it comes to driving passions and emotions. This paper examines whether the emotional trigger of sports can entail political consequences, looking at the impact of Football Club Barcelona (FCB) in Catalonia. I find that FCB performance has a strong (*temporary*) effect on Catalan identification and preferences on secession. Individuals surveyed one day after a defeat are 11% less likely to report a strong feeling of Catalan identity and are 15% less likely to support Catalan independence. By contrast, FCB titles at the end of the season increase identification and support for regional self-determination by around 11%. I find suggestive evidence that these effects can be driven by changes in euphoria, as FCB performance affects political and economic prospects as well as trust in the government.

Keywords: National identity, Political Preferences, Football, Elections.
JEL Classification: Z29, D72.

"The club is the only identity card the fan believes in. And in many cases, the team jersey, the hymn, and the flag embody endearing traditions that are expressed on the soccer fields, but come from the depths of a community's history." - Galeano (2014)

1 Introduction

Identity cleavages can be a major source of political and economic instability. Polarized societies display lower levels of trust (Alesina and Zhuravskaya, 2011), under-investments in public goods (Banerjee et al., 2005), and are more prone to conflict (Besley and Reynal-Querol, 2014). As a result, States have historically engaged in nation-building policies, fostering a common sense of identity or raising prejudice against certain ethnic minorities. Prominent examples include mass schooling and educational reforms (Bandiera et al., 2019; Cantoni et al., 2017; Clots-Figueras and Masella, 2013), propaganda and the use of media (Voigtländer and Voth, 2015; Blouin and Mukand, 2019), and even mass population re-allocations (Bazzi et al., 2019). Emotional shocks, especially those involving collective experiences, can also shape identity and values in different ways (Dell and Querubin, 2018; Depetris-Chauvin et al., 2020).

In this respect, sports have the capacity to drive people's passions and to inspire a strong feeling of belonging. Football, in particular, has often been associated with the construction of ethnic and national identities, with countless examples of clubs fostering and embracing certain identities (Foer, 2004; Shobe, 2008a,b; Úbeda-Colomer et al., 2017). Football may serve as a pseudo-religion, where players act as gods, stadiums as cathedrals, and there is a pletora of symbols such as flags and hymns (Vázquez-Montalbán, 1997). In a context of strong cleavages, this framework of rivalry, of opposition of positions, fueled by the emotional component of sports, can intensify differences between two entities in a context of strong cleavages (Ramonet, 1998). For instance, the antagonism between the GNK Dinamo Zagreb and the Red Star from Belgrade was at the epicenter of the Yugoslav Wars (Foer, 2004). Similarly, the rivalry between Glasgow Rangers and Celtic F.C. reflects a long-standing religious and political clash between Protestantism-Unionism against Catholicism-Republicanism (Úbeda-Colomer et al., 2017). Despite plenty of examples of the connection between sports, politics and identities, the extent to which they can exacerbate pre-existing cleavages and entail political consequences remains an empirical question yet to be explored.

This paper asks whether a close connection between a sports club and a feeling of identity can in-

fluence national identity, political preferences and lead to political change. I focus on the context of Catalonia and Football Club Barcelona (FCB), one of the most acclaimed clubs worldwide, and which has historically embraced and promoted Catalan identity. Leveraging survey data and information on FCB matches and trophies between 2009 and 2019, I find that FCB performance leads to sharp (*temporary*) changes in Catalan identification, preferences on decentralization and support for secession in Catalonia. My results are in line with the findings of Depetris-Chauvin et al. (2020) in the context of African football, but the implications are however the opposite: football can contribute to both unite and disunite a country. Exploiting data on FCB fans across municipalities, I find evidence that match defeats just before elections have a significant impact decreasing political participation. These results suggest that despite the temporary nature of the changes in identity and political preferences associated with football, they can translate into permanent changes by affecting electoral results.

The main data comes from the barometers of the CEO and CIS, two opinion surveys conducted regularly in Spain. The CEO barometer surveys around 1500 people in Catalonia every quarter while the CIS barometer surveys around 2500 people in Spain every month. Both surveys include an identity question in which individuals have to describe themselves as "Only Spanish", "More Spanish than Catalan ", "As Spanish as Catalan", "More Catalan than Spanish" or "Only Catalan ". They also ask individuals about their preferences regarding State decentralization and the possibility of regions to become independent. I match the survey data with detailed information on FCB matches and trophies based on the date of the interviews. Given the differences between the CEO and CIS surveys, I use them to conduct two separate but complementary analyses: one focusing on the short-run impact of FCB matches and another looking at the short-run effect of FCB titles at the end of the season.

I first assess the casual impact FCB matches on identity and political preferences exploiting the fine coverage of the CEO barometer. I compare Catalans interviewed one day before a match to those interviewed one day after, depending on the outcome of that match. I find that FCB defeats have a large effect reducing Catalan identification by 5.8 percentage points on average (-11%). Wins or ties do not seem to have any impact on identity. Note that I do not rely on the concept of unexpected or "upset losses" as in Card and Dahl (2011) because FCB was one of the best teams worldwide during this time and hence expected to win every match. Moreover, changes in iden-

tity due to FCB defeats also come hand in hand with changes in political preferences. FCB defeats reduce the share of individuals favoring more autonomy by 5.2 percentage points (-7.4%), and lowers support for secession by 7.6 percentage points (-14.6%). I explore the potential heterogeneity of the impact of FCB games and find that neither level of education or the Catalan origin of individuals seems to matter.

To make sure that the former analysis truly captures the causal effect of FCB matches, I perform a number of robustness checks. First, I verify that there are no changes in the characteristics of individuals taking part in the survey before and after matches. Second, to account for the fact that defeats could be more likely to occur in a given moment of the year or the week, I introduce month and day of the week fixed effects, which do not affect the results. Third, I perform two falsification tests, one consisting of using the outcome of the immediately next (or previous) FCB match and a second one relying on the outcomes of other teams. Both exercises yield insignificant estimates clustered around zero. Finally, I conduct a placebo exercise looking at the potential effect on less malleable characteristics, such as individuals' origin and the language spoken at home. FCB defeats are unrelated to the share of individuals who declare being Catalan or speaking Catalan.

I then assess the causal impact of FCB titles on identity and preferences exploiting the national and monthly coverage of the CIS barometer. I do so by comparing the answers of Catalan to non-Catalan depending on the number of major titles won by FCB each season while taking into account persistent regional differences and year-specific effects. I find that each additional FCB title increases Catalan identification among men by 5.2 percentage points, relative to a baseline of 45.1% percent. The effect is very similar in size to the estimates reported by Depetris-Chauvin et al. (2020) when national teams win matches in the African cup although in their case the baseline levels of ethnic identification are much lower (e.g., around 15%). It is also similar in magnitude to the effect of FCB matches but with the opposite sign. This result is not surprising given that defeats often directly compromise the chances of winning that competition. Contrarily to the case of match defeats, only men seem to react to FCB titles. As before, changes in identity among men due to FCB titles translate into sharp changes in political preferences. Each additional title is associated with an increase in the share of Catalan men favoring higher autonomy by 3.2 percentage points (+7.4%), and the right for self-determination by 4.4 percentage points (+11.7%). As a comparison, a major educational reform introducing a new curriculum and Catalan language in

schools increased Catalan identity by 2.5 percentage points, and support for self-determination by 2 percentage points for every year of exposure (Clots-Figueras and Masella, 2013).

The impact of FCB titles on political preferences lasts for a month, slightly more than the changes observed on identity. Past research shows that the impact of sports events can vary in their persistence depending on the outcome. From a couple of hours in the case of crime or domestic violence (Munyo and Rossi, 2013; Card and Dahl, 2011), to one week in judicial sentencing (Eren and Mocan, 2018) to even several months in the case of conflict (Depetris-Chauvin et al., 2020). These differences may be explained by the importance of the sport, the salience of the match, and the mechanisms involved in each case, as I discuss later.

The analysis of the impact of FCB titles relies on the assumption that FCB performance in a given year is unrelated to any other factor that could influence the outcomes of interest. I carry out a number of checks that support this claim. First, the characteristics of individuals are balanced across years, suggesting that selection on observables plays no role. Second, I show that the estimates are large and significant only in June, precisely when the season ends. Third, I show that the effects are not driven by a general change in regional attitudes, as the results remain very similar when comparing Catalan to other regions with a strong sense of identity (i.e., Basque Country and Galicia). Forth, I implement various falsification exercises using FCB titles in other years and the performance of other teams, which yield reassuring results. Finally, I perform two sensitivity analysis excluding regions and years sequentially, changes that barely affect the baseline estimates.

After establishing that FCB performance affects identity and political preferences, I explore its potential political consequences. To do so, I construct a database of Catalan municipalities combining data on all elections since 1977, FCB fan clubs across municipalities, and detailed information on matches played around the election dates. The empirical approach compares voting outcomes within and across municipalities, depending on whether there was a match just before elections and the intensity of FCB support. This analysis shows that, in recent decades, FCB matches played one day before the election decrease participation by 0.2 percentage points for every FCB fan club in the district (being the median around 10). I do not find however any impact on the vote-share of parties with a stronger identity component (i.e., ERC, CIU).

Note that the identification of a causal effect does not require FCB support across municipalities

to be exogenous since the model performs a within-municipality comparison. Moreover, elections always take place on Sundays, whether or not there was a match has thus no relation with the election date. Yet, there could still be a concern that elections with a close match may take place in a particular time of the year, and that this timing interacts with other confounding of FCB support across municipalities. Similarly, in the case of the impact of FCB defeats, the identifying variation comes only from 10 losses out of 55 matches, hence one may wonder whether by chance there could have been other shocks in those particular years with a differential impact by the intensity of FCB support. To account for potential time confounders I introduce month fixed effects as well as interact year dummies with a wide range of fixed municipality-level controls. I also implement an instrumental variable strategy to obtain exogenous variations in FCB support by exploiting satellite data on football fields across municipalities.

I conduct some additional analysis that shed some light on the mechanisms driving the impact of FCB on identity and political outcomes. Since many fans gather in pubs and fan clubs to watch matches, I use data on pubs per capita and of FCB fan clubs ("peñas") across Catalan districts to explore the role of socialization. I find that the impact of defeats is larger in areas with more FCB supporters and with a higher density of pubs and that both forces reinforce each other. This result supports the idea that shared experiences are essential drivers of changes in identity. [More discussion.] Similarly, the estimates are significantly larger in more important matches (e.g., playing at home, against rival teams, or qualifiers), speaking to the idea that defeats may trigger an emotional response (Healy et al., 2010; Passarelli and Tabellini, 2017; Van Winden, 2015) . In fact, I do find evidence that FCB titles affect euphoria, manifested by more positive views on the political and economic situation, as well as trust in Government. These findings contrast with the results of Depetris-Chauvin et al. (2020) in the context of African national football, where they find no impact of matches on well-being, economic prospects, or trust in Government. They align however with the work of Healy et al. (2010) which reports changes in mood after American football matches.

This paper contributes to different strands of the literature. First, it relates to the work on the formation of national and ethnic identities. Most of this research has focused on analyzing the impact of nation-building policies (Bazzi et al., 2019; Bandiera et al., 2019) . In some cases, these policies have led to a backlash and a reinforcement of ethnic identities (Fouka, 2020; Dehdari and

Gehring, 2019). Some other work has examined the effect of emotional experiences caused by war (Dell and Querubin, 2018), natural disasters (**?**) or even sports successes (Depetris-Chauvin et al., 2020). In a closely related work, Depetris-Chauvin et al. (2020) find that victories in the African football cup can increase national identity and inter-ethnic trust, and even reduce the intensity of conflict. My results complement their findings and bring important new insights. On one hand, sports can serve to either foster unity or polarization, depending on the identity or ideas they portrait.

My results provide new evidence on how random shocks can affect voting behavior. For instance, previous work has documented the role of fatigue (Augenblick and Nicholson, 2016), past life satisfaction (Liberini et al., 2017), and weather shocks (Lind, 2019; Arnold and Freier, 2016; Fujiwara et al., 2016; Meier et al., 2019) affecting electoral participation and support for the status quo. In a related work, Healy et al. (2010) find that unexpected wins in local college football increase the vote-share for the incumbent, regardless of its party affiliation. My results show that sporting events may systematically benefit certain political parties, as long as there is a connection with the values and identity the team or the fans portray. This finding has strong policy implications, as it suggests that politicians or the government could strategically exploit sports teams for their own advantage. Similarly, it brings attention to the validity of high-stakes decisions adopted by a short margin (e.g., the Brexit referendum), to the extent that they can be influenced by arguably irrelevant factors. On top of that, if short-lived changes in preferences affect voting outcomes, they can alter the politicians who are elected, and in turn, the policies and institutions that will shape the future of that society.

I also contribute to the literature on the determinants of political preferences and attitudes, and the way in which they change in the short run. As shown by Clots-Figueras and Masella (2013) and Rodon and Guinjoan (2018), a stronger feeling of regional identity tends to be associated with favorable views on decentralization and secession. My findings show that these preferences are highly volatile and they may be influenced by sporting events. Moreover, political preferences seem more malleable than identity and can change even if the latter remains constant. It is crucial to highlight that the identities and values a club portray are central to the effects one may expect. In fact, football could be used as a platform to reduce prejudice between conflicting groups in certain settings (Mousa, 2020).

Finally, my results bring new evidence on the major consequences of sporting events, and in particular, the performance of local teams. Previous research has shown that, across a variety of sports, the performance of local teams can affect crime (Munyo and Rossi, 2013), domestic violence (Card and Dahl, 2011), students' achievement (Lindo et al., 2012), judicial decisions (Eren and Mocan, 2018) and even infant health (Duncan et al., 2017) . Several of these studies point to emotions as a trigger of changes in behavior (Van Winden, 2015; Passarelli and Tabellini, 2017). I provide additional evidence that football affects euphoria, which in turn may shape individuals' beliefs about themselves and their identity group, and having ultimately political consequences.

2 Background: Sports, politics and Catalonia

2.1 FCB, Catalan identity and politics

Most Catalan are football enthusiasts and fierce supporters of the FCB. According to survey data (CIS, 2014), 74% of adults in Catalonia sympathized with a football team and 78% of those are FCB fans. While the support for FCB is similar between men and women, there is a significant gap in the share who declared being interested in football, 79% and 68% respectively.

On average, FCB fans hold a strong Catalan identity, favor the right of self-determination, and vote for pro-independence parties. In terms of identity, 52% of FCB fans consider themselves only Catalan or more Catalan than Spanish (42% in Catalonia). Yet, there is a substantial share of fans with a dual identity, 35% of them identify equally with Catalonia as with Spain. As of political preferences, over 51% of fans support the right of self-determination (43% in Catalonia) and 49% voted for one of the two main pro-independence parties (CiU, ERC) in the past elections (37% in Catalonia). Note that the distribution of identity and political preferences is more concentrated among FCB fans than in Catalonia. For instance, only 20% feel less Catalan than Spanish (27% in Catalonia) and only 17% supports a centralized state (24% in Catalonia). Similarly, the support for traditional parties (PP, PSOE) is lower among FCB fans than in Catalonia, 37% compared to 41%.

The FCB stadium has served as a loudspeaker of Catalan identity and pro-independence views, especially in recent years. For example, the stadium is often riddled with *esteladas* (i.e., the independence flag) and other pro-independence messages. On several occasions, and with the permission of the club, giant placards were displayed with messages such as "Free Catalonia´´, "Welcome to the Catalan Republic´´, "Self-determination is a right, not a crime´´. [1] In some matches, *esteladas* and political leaflets were handled at the entry (REF). Since 2010, in the minute 17:14 of every match fans start to shout *"Independencia ´´*, commemorating the date (1714) in which Catalan troops lost in the war of the Spanish Succession. Moreover, in the past decade, most fans react with whistles and boos to Spanish symbols such as the anthem or the visits of the Spanish king. Given how Catalan identity and political views manifest in the stadium, sharing them collectively can become a central element of being an FCB fan and incentive individuals to embrace them.

---

[1]The UEFA (The Union of European Football Associations), in fact, fined the club for those incidents, as the display of political symbols is not allowed in stadiums (REF).

The connection between the FCB and Catalan identity goes far beyond the fans, the direction of the cub has historically contributed to strengthening such link, embracing and promoting Catalan identity. Already in the 1910s and 1920s, the club took a stand demanding higher autonomy for the region, used the Catalan language in their meetings and documents, etc. (Tomás and Porta, 2016). It is worth pointing out that the president at the time, Gaspar Rosés, had been the leader of the Catalan nationalist party Liga Regionalista. As an illustration, the last assembly of 1920, the minutes reflected *"We are from FCB because we are from Catalonia. We do sports because we build a nation"* (Tomás and Porta, 2016). In the last two decades, amidst growing political tension in Catalonia, the FCB direction and stuff openly expressed their political views and support for an independent Catalonia. For instance, the club offered the stadium for different pro-independence events and as a voting poll for an independence referendum (deemed illegal by the Constitutional Court) to be held in 2011. It also released official statements favoring the right of self-determination and condemning the sentence of the Supreme Court regarding the incidents around the independence referendum of 2017 (which had previously been deemed illegal by the Constitutional court). Members of the stuff of the club have also been engaged with the Catalan independence movement. The former president Joan Laporta (2003-2010) founded a pro-independence party himself while former coach Pep Guardiola (2010-2014) symbolically joined the independence coalition *Junts pel Sí* in the 2015 elections. Several players also manifested their political views in favor of higher autonomy or even secession repeatedly. For instance, during the (illegal) referendum of 2017 Gerard Pique posted on his Twitter account a picture of him voting and encouraging others to do so: *"I have already voted. Together we are unstoppable defending democracy"*. This tweet was then *'liked'* by over a hundred thousand people.

3 Data

3.1 Football data

I collect information on all FCB matches (first-team, men) over the 2012-2020 period. These data include the date of the match, location, competition, opponent, and final score. I link the matches data to individuals in the survey data based on the date of the match (further details in the next section). I also use information on the titles won every season and the location of all FCB fan clubs. Additionally, I use information on other teams (e.g., Real Madrid) for robustness checks.[2]

---

[2]See the official FCB site for matches results, titles per season and the location of fan clubs. Similar matches' data on other teams can be found in several websites, for instance, `http://www.worldfootball.net/`. Real Madrid titles are

In the period under study (2009-2020), FCB won 17 major titles, averaging 1.55 every year (8 Spanish leagues and 6 Spanish cups, 3 UEFA Champions League). After linking the matches data with the individual survey data, the final sample comprises 55 matches, among which 41 victories, 4 ties and 10 defeats. Across competitions 32 corresponds to the Spanish league, 9 to the Spanish cup and 14 to the UEFA Champions League. Table B.1 in the Appendix provides further details about the matches in the sample.

3.2 Individual survey data

I use individual survey data from two representative opinion surveys conducted regularly in Spain ("Barómetro", henceforth BAR, elaborated by the CIS[3]) and in Catalonia ("Baròmetre d'Opinió Polític", henceforth BOP, elaborated by the CEO[4]). The BAR interviews around 2,500 people every month across all Spanish regions while the BOP interviews around 2,000 people every four months only in Catalonia. The objective of both surveys is to measure the opinions of the population regarding politics, the economy and different social issues.[5] Interviews are conducted either in Spanish or the regional language (e.g., Catalan, Galician).

I focus on three main outcomes relating to the degree of regional identity, preferences on state decentralization and regional independence. The main outcome variable is a measured of identity based on the question: *"How do you identify yourself? I feel..."* after which respondents pick one of the following: (1) *Only Spanish*, (2) *More Spanish than [region demonym]*, (3) *As Spanish as [region demonym]*, (4) *More [region demonym] than Spanish*, (5) *Only [region demonym]*. I construct the variable *Identity* as a dummy equal to 1 if individuals identify more strongly with their region than with Spain (i.e., answers 4 and 5) and zero otherwise. Regarding preferences on decentralization, in the BAR survey I rely on the question *"What kind of organization for the State you would favor?"* with answers: (1) *Centralized State*, (2) *Regions with less autonomy*, (3) *As of today*, (4) *Regions with more autonomy*, (5) *Regions that could become independent States*. I construct the variable *Autonomy* as a dummy equal to 1 if individuals favor higher regional autonomy (i.e., answers 4 and 5) and

---

retrieved from their official site and the league placement of other teams from `http://www.marcadores.com/futbol/espana/liga-bbva/clasificaciones.html`. The period covered (2012-today) comes from the availability of identity and political preferences questions in the individual survey data.

[3]Centro de Investigaciones Sociológicas.

[4]Centre d'Etudis d'Opinó.

[5]These questions are very similar to those included in the World Values Surveys or the Afrobarometer.

zero otherwise, and the variable *Independence* as a dummy equal to 1 if individuals favor regional self-determination (i.e., answer 5). Note that depending on the survey (BAR or BOP), the questions may differ slightly in their wording but capture the same idea. The BOP survey, for instance, includes a yes/no question on whether Catalonia should become an independent State. Table X in the Appendix details these questions and how the outcome variables are defined. I also use data on voting intentions (e.g., which party they intend to vote in the next elections) and on a wide range of demographic characteristics.

Figure A.1 and A.2 illustrate the CIS survey data about regional identity and preferences on state decentralization and independence for three regions in Spain: Catalonia, Galicia, and Madrid. Regional identity is particularly strong in Catalonia, with 43% of the population "feeling" more Catalan than Spanish, compared to 23% in Galicia and 5% in Madrid. A large share of the population holds a balance dual-identity, meaning that they identify equally with their region as well as with Spain: 40% in Catalonia, 69% in Galicia, and 43% in Madrid. On the other side, 15% of the population in Catalonia holds a strong national identity, compared to 8% in Galicia and 43% in Madrid. There is a parallel between regional identity and preferences on state decentralization. Regions with stronger identities tend to be more favorable to higher regional autonomy while those with more nationalistic views are more inclined towards less autonomy or a centralized State. As shown in Figure A.2, 64% of the population in Catalonia favors more autonomy, compared to 15% in Galicia and 17% in Madrid. The current level of autonomy is preferred by 22% in Catalonia, 60% in Galicia, and 27% in Madrid while a centralized State is preferred by 10% of the population in Catalonia, 15% in Galicia and 32% in Madrid. In terms of the evolution in recent years, regional identity feelings and preferences on state decentralization have remained fairly stable in the period under study.

To validate the results at a finer level and to explore in depth certain mechanisms I use additional survey data from the city of Barcelona (Cristancho et al., 2020). This survey was conducted on 1500 residents of the city between May and June 2015 and contains information on the neighborhood of residence, Catalan identification and support for independence. Moreover, it also asks respondents about the preferences over secession of a number of close contacts (e.g., family and friends). The timing of the survey is ideal since the interviews were conducted during weeks in which FCB succeeded in winning 2 titles and there were also local elections.

3.3 Elections data

To assess the impact of FCB performance on electoral results I use municipality-level data of all the elections held between 1977 and 2020.In particular, I use participation rates and the vote-shares of different parties across the 947 Catalan municipalities. Given that some parties disappear and new ones appear over time, I group them in two blocks, those (strongly) promoting Catalan identity and pursuing more self-government and those more inclined towards the status quo. Section X in the Appendix provides further details on the political parties and how they are classified into identity-based or not. Overall, I gather information about 46 elections, 11 municipal, 12 regional, 15 national, and 8 European. Additionally, I use data on the participation and results of the popular referendums about the Catalan independence of 2014 and 2017.[6]. Table B.2 in the Appendix provides more details on the elections and FCB matches played close them.

4 Impact of FCB matches on national identity and political preferences

4.1 Empirical Strategy

The first analysis examines the short-run impact of FCB matches on Catalans' identity and political preferences. I use survey data from the CEO barometer between 2014 and 2018, the period in which the date of the interview was recorded. The research design exploits variation in the timing and outcome of matches, comparing the answers of individuals interviewed one day after a match rather before and depending on its result. The estimating equation is:

$$y_{ig}^{BOP} = \theta_g + \rho \, PostDefeat_{ig} + \psi \, PostDraw_{ig} + \phi \, PostVictory_{ig} + u_{ig} \, , \qquad (1)$$

where i indexes individuals and g indexes matches, θ_g represents match fixed effects, $PostDefeat_{ig}$, $PostDraw_{ig}$, and $PostVictory_{ig}$ are dummy variables equal to one if individual i is interviewed after FCB lost/tied/won match g and zero if interviewed the day before the match, and u_{ig} are standard errors clustered at the district and match level. I focus on similar outcome variables as in the analysis of FCB titles, with the addition of having explicit information about support for independence. The questions of identity and preferences on autonomy in the BOP surveys have almost the same wording as in those of the CIS. The outcome of identity is a dummy equal to one

---

[6]These data can be retrieved from http://www.infoelectoral.mir.es/infoelectoral/min/areaDescarga.html?method=inicio.

if individuals declare feeling more Catalan than Spanish or only Catalan and zero otherwise. The outcome of preferences for autonomy is a dummy equal to one if individuals think that Catalonia has insufficient self-government. Finally, the third outcome is a dummy reflecting support for Catalan independence (i.e., "I want Catalonia to become an independent state"). The coefficients ρ, ψ and ϕ measure the gap in outcomes between individuals interview one day before or after a match depending on whether FCB lost, tied or won that match.

To interpret the estimates of equation (1) as one needs a series of assumptions. First, being interviewed just before or after a match should be as good as random, in other words, the characteristics of respondents should be balanced. As shown in Table 1 Panel A, the timing of the interview is not correlated with a wide range of individual characteristics. Similarly, respondents should not be systematically different in the days surrounding an FCB defeat or win. Panel B largely supports this assumption. With the exception of college education, there are no big differences in the characteristics of individuals interviewed around matches depending on its outcome. The same conclusion holds true when examining how the pre-post match difference in the variables may correlate with the outcome of the match (i.e., double difference). As shown in Panel C, the only characteristic that is unbalanced is the employment status, although this result does not come with any difference in terms of college education or SES status. Although I will control for these variables in the analysis, their inclusion has literally no influence on the results. Importantly, FCB matches do not take place on the same day of the week. Most matches from the League competition take place on Saturday and Sunday, while those from the Spanish Cup and Champions League competitions mostly on Tuesdays and Wednesdays. In a few instances, FCB has also played matches of these competitions in other days of the week.

Table 1: Balance of FCB matches and individuals' characteristics

Covariate	N	Mean	SD	Panel A: Post-match		Panel B: Defeat		Panel C: Post-defeat	
				Estimate	Std. Errors	Estimate	Std. Errors	Estimate	Std. Errors
Men	8,400	0.50	0.50	-0.012	0.011	0.007	0.015	0.016	0.028
Senior	8,400	0.47	0.50	-0.009	0.016	-0.024	0.021	-0.032	0.040
Couple	8,400	0.64	0.48	-0.014	0.015	0.013	0.021	-0.018	0.039
Children	8,400	1.27	1.21	0.012	0.039	-0.018	0.052	0.012	0.098
College	8,400	0.41	0.49	0.005	0.016	-0.052***	0.021	-0.017	0.040
Employed	8,400	0.54	0.50	0.009	0.016	0.034	0.021	-0.076**	0.040
Low SES	8,400	0.38	0.48	-0.003	0.015	0.009	0.021	0.021	0.039
Catalan born	8,400	0.73	0.44	0.001	0.014	0.003	0.019	-0.047	0.036
Speaks Catalan	8,400	0.49	0.50	-0.020	0.016	-0.029	0.021	0.007	0.040

Note: This table shows the relationship between individuals' characteristics and the date of the interview. Panel A compares individuals before/after a match, Panel B individuals around a defeat vs. a tie or win, and Panel C how the gap before/after changes depending on a defeat (i.e., double difference). Each Panel-row represents a separate regression. For instance, the first row in Panel A shows that the share interviews who are male is 1.2% lower the day after a match than the day before. The standard errors are clustered at the *match* level. *, **, *** indicate p-values below 10%, 5%, an 1% respectively.

4.2 Results: FCB matches, national identity and political preferences

MAIN RESULTS: Figure 1 illustrates the evolution of Catalan identity (Figure 1.a) and preferences on regional autonomy and Catalan independence (1.b and 1.c) around the days of a FCB match. More precisely, it plots the short-run impact of of a FCB defeat (i.e., OLS estimates of the interaction between day fixed effects and a defeat dummy). As shown in the three figures, FCB defeats lead to a sudden fall in Catalan identification, a drop in preferences on higher regional autonomy and a decrease in support for Catalan independence. These effects seem to last only for one day as the outcomes return to their baseline levels immediately after.

Table 2 summarizes the short-run effect of FCB matches on identity (Panel A) and political preferences (Panels B and C). It presents the results of estimating equation (1) by OLS relying on the sample of individuals interviewed one day before and after a match. The main message of the table is that FCB defeats have a large impact lowering Catalan identification (-11%) and reducing preferences on decentralization (-7%) and support for Catalan independence (-15%).

Figure 1: Impact of FCB defeats on Catalans' identity and political preferences

Figure 1.a: Strong Catalan identity

Figure 1.b: Favors higher autonomy

Figure 1.c: Supports Catalan independence
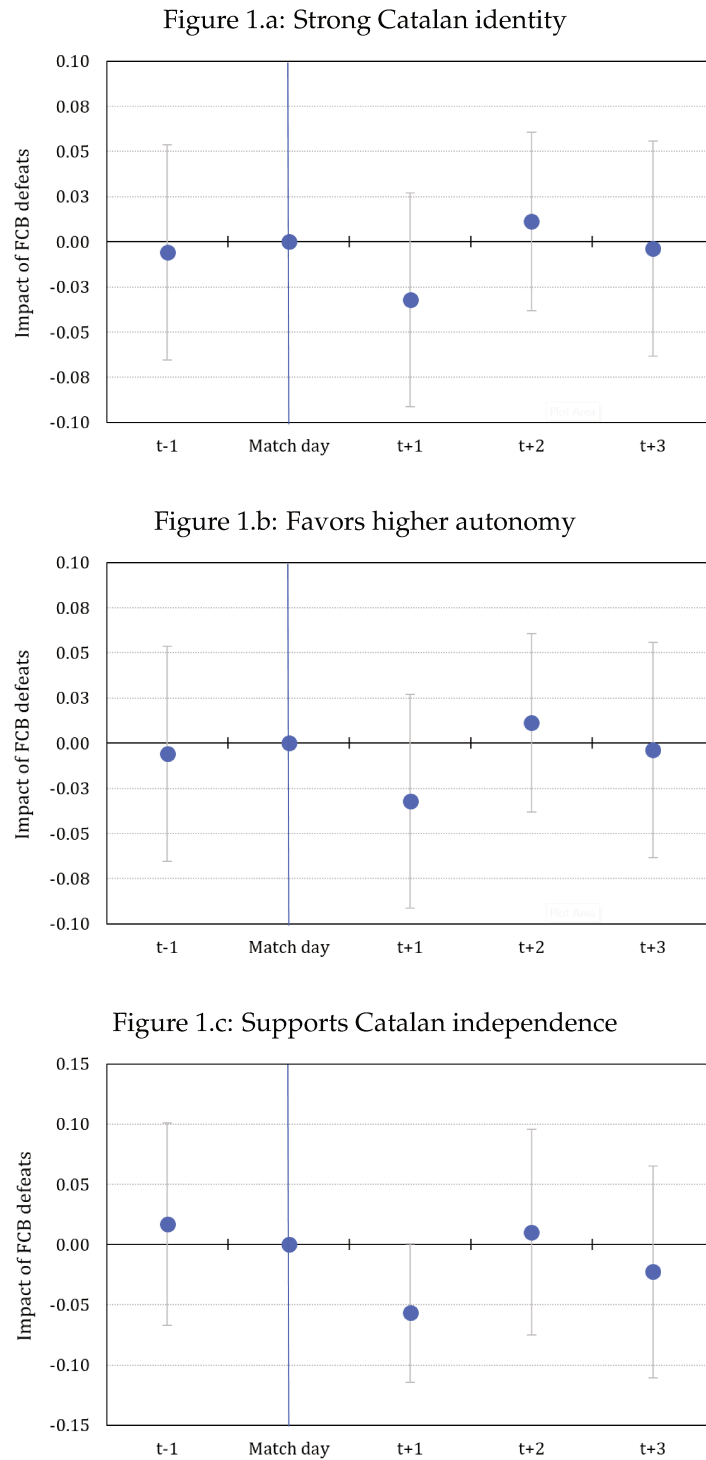
Note: This figure plots the impact of FCB defeats on Catalans' identity and political preferences over time (i.e., the results of estimating equation (1) interacting the defeat dummy with day dummies). The outcomes are dummies that represent identifying more with Catalonia than with Spain (Panel A), favoring higher regional autonomy (Panel B) and supporting Catalan independence (Panel C). As an illustration, Panel C shows that one the day after match individuals are around 6 percentage points less likely to support Catalan independence if FCB lost rather than won. Dotted lines represent 90% confidence intervals based on standard errors clustered at the *match* level.

Table 2: Impact of FCB matches on identity and political preferences

	Panel A: Strong Catalan identity					
	1	2	3	4	5	6
Post-Defeat	-0.033	-0.026	-0.047	-0.058*	-0.068	-0.053*
	0.047	0.048	0.038	0.033	0.047	0.029
Post-match	0.002	0.003	-0.012			
	0.016	0.017	0.019			
Defeat	0.010	0.009				
	0.042	0.042				
Post-Draw				0.008	0.033	-0.016
				0.080	0.058	0.096
Post-Victory				-0.014	-0.017	-0.012
				0.019	0.025	0.018
	Panel B: Favors higher autonomy					
Post-Defeat	-0.029	-0.026	-0.035	-0.052**	-0.058	-0.048**
	0.040	0.040	0.028	0.024	0.049	0.022
Post-match	-0.003	-0.002	-0.016			
	0.015	0.015	0.014			
Defeat	0.017	0.018				
	0.026	0.025				
Post-Draw				-0.048***	-0.098***	0.001
				0.011	0.013	0.005
Post-Victory				-0.013	-0.01	-0.02
				0.016	0.020	0.019
	Panel C: Supports Catalan independence					
Post-Defeat	-0.059	-0.057	-0.074**	-0.076***	-0.058*	-0.103***
	0.038	0.038	0.034	0.026	0.033	0.026
Post-match	0.008	0.010	-0.002			
	0.019	0.019	0.022			
Defeat	0.037	0.036				
	0.034	0.034				
Post-Draw				0.016	0.019	0.02
				0.102	0.051	0.141
Post-Victory				-0.004	-0.001	-0.01
				0.021	0.027	0.02
Observations	7,743	7,726	7,726	7,726	3,881	3,845
Individual controls	No	Yes	Yes	Yes	Yes	Yes
Match fixed effects	No	No	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Men	Women

Note: This table summarizes the results on the impact of FCB matches on Catalan's identity and political preferences. The outcome of identity represents feeling more Catalan than Spanish or just Catalan. Support for Catalan independence measures whether individuals support that Catalonia becomes an independent State. Post-Defeat/Draw/Victory are dummies indicating being interview the day after such outcome rather than the day before. Controls include dummies for male, older than 50, married, children, college education, employment and low SES. Standard errors clustered at the match level.*, **, *** indicate p-values below 10%, 5%, an 1% respectively.

Starting by Panel A in Table 2, the first specification in column (1) includes three dummies that capture (i) being interviewed around a defeat (Defeat), (ii) being interviewed the day after a match, and (iii) being interviewed the day after a defeat (Post-Defeat). None of these estimates is however precisely estimated, with standard errors being larger than the coefficients themselves. Adding individual controls in column (2) reduces slightly the point estimates but remain imprecise. Column (3) includes match fixed effects which effectively means comparing individuals around the same match and forces the dummy Defeat to be excluded. In this case, the point estimate of the interaction term is larger and the standard errors considerably lower (p-value = 0.21).

The preferred specification in column (4) includes match fixed effects and dummies for being interviewed one day after each potential outcome of a match (i.e., victory, tie or defeat). The result indicates that those interviewed after a defeat are 5.8 percentage points less likely to report a strong Catalan identity, rather than those interviewed just before that match (i.e., an 11% decrease). The estimates for the other match outcomes are imprecisely estimated. Columns (5) and (6) split the sample between men and women. The effect of a defeat is not significantly different across sexes, while the point estimate is larger among men the standard errors are relatively higher. This finding contrasts with the results in the previous section in which women did not seem to react to FCB titles. One possibility is that the emotional responses to major triumphs and crucial defeats operate differently for men and women.

The results show that FCB defeats not only reduce Catalan identity but lower support for decentralization and Catalan independence. Column (4) in Panel B shows that FCB defeats reduce the share of Catalan thinking Catalonia has insufficient autonomy by around 5.2 percentage points (7.4% decrease). The point estimate is slightly larger for men but less precisely estimated. Column (4) in Panel C shows that FCB defeats decrease support for Catalan independence by 7.6 percentage points (14.6% decrease). Interestingly, the point estimate is larger for women than men in this case. One hypothesis for this result is that defeats may increase intimate partner violence (Card and Dahl, 2011) which in turn spill over secession preferences of women . There is evidence indeed pointing that FCB upset losses increase domestic violence right after the match (Montolio and Planells-Struse, 2016).

4.3 Additional results: FCB matches, shared experiences and heterogeneity

In this section, I explore whether the impact of FCB matches may vary depending on the type of match and the environment (Table 3) or the characteristics of individuals (Table 4).

Columns (1) and (2) of Table 3 show that the impact of home and away matches differ. Only FCB defeats away seem to decrease Catalan identity and support for independence. This evidence however should be taken with caution given that FCB only lost 2 out of the 26 home matches in the sample (vs. 7/26 away). Next, I examine the impact of high-stakes matches and against traditional rivals.[7] As shown in columns (3) and (4), draws in challenging matches or against traditional rivals can increase Catalan identification and support for independence. This result may suggest that when the team is at the brink of losing/ winning a crucial match, emotions emerge more strongly boosting identity and preferences. Interestingly, Champions League defeats decrease identification and preferences for secession while defeats against traditional rivals do not. Another question is whether social interactions amplify or reduce these effects. Most supporters watch the matches with their friends in pubs or fan clubs. To explore this, I use data on the FCB fan club (*peñas*) across district and pubs per capita. Columns (5) and (6) show that the impact of FCB defeats is entirely driven by areas with fan clubs and pubs density above the median.

Table 4 summarizes the results interacting the post-defeat dummy with different individual traits.The interaction term in all regressions is imprecisely estimated although the estimates are often large, hence the results should be taken with a grain of salt. Column (1) shows that men seem more responsive than women in terms of identity while the opposite is true regarding support for independence. While college education does not seem to matter (column 2) individuals from lower SES seem more affected by defeats (column 3). The impact of defeats on identity seems lower for individuals regularly speaking both Catalan and Spanish (column 4) and born outside Catalonia (column 5). There is however no apparent difference in the impact on supports for independence or if any, those born outside Catalonia might react more strongly. Note that these individuals have on average lower rates of Catalan identification and support for independence.[8] Finally, columns (6) focus on the sub-sample of individuals born outside Catalonia. FCB defeats do not seem to have a differential impact depending on how many years they have spent in the region.

---

[7]Instead of restricting only to Real Madrid, I consider also Atlético de Madrid and Espanyol, which gives a total of 9 matches in the sample, among which 5 wins, 3 defeats and 1 draw.

[8]Clots-Figueras and Masella (2013) show that the impact of a major educational reform on Catalan identity and secession preferences was similar among children of Catalan and non-Catalan parents.

Table 3: Impact of FCB matches depending on type of match and location

	Panel A: Strong Catalan identity					
	1	2	3	4	5	6
Post-Defeat	0.023***	-0.071*	-0.098***	-0.013	0.0005	0.005
	0.004	0.036	0.014	0.093	0.042	0.029
Post-Defeat x Interaction					-0.152***	-0.132***
					0.047	0.043
Post-Draw	0.110***	-0.117***	0.109***	0.107***	0.008	0.008
	0.003	0.001	0.002	0.004	0.08	0.08
Post-Victory	-0.036	0.018	-0.017	-0.026	-0.014	-0.014
	0.026	0.022	0.03	0.064	0.019	0.019
	Panel B: Favors higher autonomy					
Post-Defeat	0.094***	-0.072***	-0.055*	-0.024***	-0.010	0.016
	0.004	0.016	0.029	0.009	0.024	0.025
Post-Defeat x Interaction					-0.109**	-0.140***
					0.044	0.040
Post-Draw	-0.033***	-0.064***	-0.035***	-0.030***	-0.048***	-0.048***
	0.002	0.001	0.003	0.005	0.011	0.011
Post-Victory	-0.023	-0.001	-0.028	0.092***	-0.013	-0.013
	0.024	0.015	0.024	0.031	0.016	0.016
	Panel C: Supports Catalan independence					
Post-Defeat	0.029***	-0.091***	-0.112***	-0.034	-0.023	-0.042
	0.005	0.025	0.013	0.053	0.035	0.028
Post-Defeat x Interaction					-0.140***	-0.072**
					0.047	0.031
Post-Draw	0.146***	-0.141***	0.145***	0.144***	0.016	0.016
	0.002	0.001	0.002	0.004	0.102	0.102
Post-Victory	-0.031	0.034*	-0.012	0.034***	-0.004	-0.004
	0.031	0.019	0.032	0.009	0.021	0.021
Observations	3,841	4,014	3,000	1,109	8,066	8,066
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Match fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Interaction term					Fans density	Bars density
Sample	Home	Away	Champions	Rivals	Full	Full

Note: This table summarizes the impact of FCB matches on Catalan's identity and preferences depending on characteristics of the match or the location. Post-Defeat/Draw/Victory is a dummy indicating being interview the day after a defeat/draw/victory rather than the day before. Controls include dummies for male, older than 50, married, children, college education, employment and low SES. Standard errors clustered at the match level.*, **, *** indicate p-values below 10%, 5%, an 1% respectively.

Table 4: Impact of FCB matches depending on individual's characteristics

	Panel A: Strong Catalan identity					
	1	2	3	4	5	6
Post-Defeat	-0.038	-0.066**	-0.037	-0.063**	-0.043	-0.073
	0.033	0.031	0.031	0.029	0.028	0.077
Post-Defeat x Interaction	-0.041	0.023	-0.052	0.030	0.036	0.0002
	0.036	0.04	0.039	0.049	0.042	0.002
Post-Draw	0.008	0.008	0.008	0.012	-0.006	-0.089
	0.046	0.046	0.046	0.045	0.043	0.072
Post-Victory	-0.014	-0.014	-0.014	-0.013	0.002	0.005
	0.014	0.014	0.014	0.014	0.014	0.023
Observations	8066	8066	8066	8066	8066	2192
R2	0.07	0.07	0.07	0.078	0.182	0.076
	Panel B: Favors higher autonomy					
Post-Defeat	-0.056*	-0.048	-0.060**	-0.050*	-0.025	-0.041
	0.031	0.029	0.030	0.027	0.027	0.092
Post-Defeat x Interaction	0.008	-0.009	0.021	-0.007	-0.039	-0.002
	0.034	0.037	0.037	0.048	0.042	0.002
Post-Draw	-0.048	-0.048	-0.048	-0.047	-0.059	-0.015
	0.043	0.043	0.043	0.043	0.042	0.089
Post-Victory	-0.013	-0.013	-0.013	-0.013	-0.003	0.017
	0.014	0.014	0.014	0.014	0.013	0.029
Observations	7823	7823	7823	7823	7823	2110
R2	0.042	0.042	0.042	0.042	0.103	0.075
	Panel C: Supports Catalan independence					
Post-Defeat	-0.097***	-0.080**	-0.052	-0.079***	-0.037	-0.117
	0.034	0.032	0.032	0.029	0.028	0.077
Post-Defeat x Interaction	0.040	0.010	-0.060	0.015	-0.045	0.0003
	0.037	0.040	0.040	0.051	0.043	0.002
Post-Draw	0.016	0.016	0.016	0.021	-0.003	-0.029
	0.047	0.047	0.047	0.047	0.044	0.076
Post-Victory	-0.004	-0.004	-0.004	-0.004	0.011	0.034
	0.015	0.015	0.015	0.015	0.014	0.024
Observations	7726	7726	7726	7726	7726	2137
R2	0.06	0.06	0.06	0.068	0.171	0.065
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Match fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Interaction term	Men	College	Low SES	Both languages	Non Catalan	Arrival year
Sample	Full	Full	Full	Full	Full	Non Catalan

Note: This table summarizes the impact of FCB matches on Catalan's identity and preferences depending on respondents' characteristics. Post-Defeat/Draw/Victory is a dummy indicating being interview the day after a defeat/draw/victory rather than the day before. Controls include dummies for male, older than 50, married, children, college education, employment and low SES. Standard errors clustered at the match level.*, **, *** indicate p-values below 10%, 5%, an 1% respectively.

5 Impact of FCB titles on national identity and political preferences

5.1 Empirical Strategy

To assess the impact of FCB annual performance on national identity and political preferences I rely on variation in the number of titles won by FCB every season.[9] I use BAR survey data between 2009 and 2019, all the years for which the questions on identity and preferences are available. The main estimating equation is:

$$y_{i,r,t}^{BAR} = \theta_r + \theta_t + \delta\,\theta_r \times Year_t + \gamma\,Catalan_{i,t} \times FCBTitles_t + \epsilon_{i,r,t}\,, \qquad (2)$$

where i indexes individuals, r regions, and t years. $Catalan_{i,t}$ equals one if living in Catalonia and zero otherwise, $FCBTitles_t$ is the number of titles won by FCB in season t and $\epsilon_{i,t}$ represents standard errors clustered at the region-year level. As in the analysis on FCB matches, I focus on three different outcome variables $y_{i,t}^{BAR}$, one for identity and two for preferences (as detailed in Table X). The identity outcome represents whether individuals report feeling more identified with their region than with Spain (e.g., "I feel more Catalan than Spanish or only Catalan "). The second outcome reflects preferences for higher regional autonomy (e.g., "Autonomous regions should have more autonomy" or "Regions should be allowed to decide whether they want to become independent States"). The third outcome indicates support for regions' self-determination (e.g.,"Regions should be allowed to decide whether they want to become independent States"). The coefficient γ captures the differential impact of FCB titles on Catalan compared to the rest of Spanish. Note that the fixed effect θ_r measures the average gap in outcomes between regions, θ_t captures yearly changes affecting all regions equally, and $\theta_r \times Year_t$ represents yearly linear trends specific for each region.

To interpret the estimate of γ as causal at least three assumptions are needed. First, the characteristics of respondents to the survey should not be correlated with FCB performance. For instance, it could be that in a year that FCB performs badly some individuals decline to participate in the survey (either fully or partially), or those taking part are different in some dimension. As shown in Table 5, FCB titles are significantly correlated with some characteristics such as earnings, the probability of being married, or the likelihood of being interview on a weekend. However, when

---

[9]Note that I focus on the three major titles; the Spanish League (LaLiga), the Spanish Cup (Copa del Rey), and the UEFA Champions League. I use the survey data of July (i.e., some weeks after these competitions have finished).

comparing Catalan to the rest of Spanish most of these differences disappear. As robustness, I will shock that controlling or not for these variables has virtually no effect on the estimates. Second, FCB performance should not be affected by changes in the nationalist sentiments or political views of Catalan. This assumption could be violated if players performed better when Catalan sentiments were on the rise (e.g., because they share those sentiments, because they receive more support from the fans, because the direction/club changed some practices, etc.). I will perform a placebo exercise in which I take into account the number of Catalan players in the team and show that this has no effect. Finally, there should be no events influencing simultaneously FCB performance and the political attitudes of Catalan. This could be violated if the region government tried to foster national sentiments and, at the same time, boost FCB performance by supporting the club economically. This concern is unlikely to effectively take place given the financial independence of the club and the limited support received by either the regional or Central Government.

Table 5: Balance of FCB titles and individuals' characteristics

Covariate	N	Mean	SD	Panel A: FCB titles		Panel B: FCB titles X Catalan	
				Estimate	Std. Errors	Estimate	Std. Errors
Male	28,820	0.49	0.50	-0.002	0.001	0.003	0.003
Age	14,038	47.7	17.7	0.205	0.160	-0.226	0.346
College educated	14,038	0.39	0.49	0.001	0.006	0.006	0.005
Employed	14,038	0.50	0.50	0.010	0.006	0.003	0.010
Left ideology	14,038	0.22	0.41	0.023	0.006	-0.013	0.010
Catholic	14,038	0.61	0.49	0.017	0.008	-0.023	0.010
Married	12,824	0.51	0.50	-0.037	0.007	-0.003	0.010
Earnings	7,802	1121	823	174.0	25.48	36.64	24.33
Weekend	14,038	0.32	0.47	0.042	0.010	0.022	0.025
Morning	14,005	0.59	0.49	-0.003	0.011	0.032	0.011

Note: This table shows the relationship between FCB titles and individuals' characteristics. The sample is restricted to men. Panel A shows the relationship between different covariates and the number of FCB titles every year (including region fixed effects) while Panel B shows how this relationship changes between Catalan and not-Catalan (controlling for region and year fixed effects as in equation 1). The standard errors are clustered at the *region X year* level. *, **, *** indicate p-values below 10%, 5%, an 1% respectively.

5.2 Results: FCB titles, national identity and political preferences

MAIN RESULTS: I start by illustrating the impact of FCB titles on Catalan identity and political preferences over time. Figure 2 plots OLS estimates of a more sophisticated version of equation (2) in which the interaction ($Catalan \times FCBtitles$) is further combined with month dummies. The estimates for the months prior to June (i.e., before the end of the season) serve thus as a placebo exercise since the number of FCB titles won that year is still unknown for individuals.

Figure 2.a shows that FCB titles have a strong impact rising Catalan men's identification right after the season ends (i.e., the first week of June), but such effect dissipates over time and fades completely by mid-July.[10] Figure 2.c shows a similar pattern in which FCB titles boost men preferences for regional self-determination, with the estimates becoming large in June and early July before going back to zero by mid-July. The impact on preferences seems to decay more slowly than the one on identity, and also to kick in slightly earlier. Figure 2.b depicts a less clear picture, with the estimated coefficient only significantly different than zero in the second week of June. Note that this outcome combines individuals favoring both (i) more autonomy and (ii) the right of self-determination, and hence, it will not change if individuals shift their views from the former to the latter.

Table 6 summarizes the results on the short-run impact of FCB titles on identity (Panel A) and political preferences (Panels B and C) by presenting OLS estimates of equation (2) based on individuals interviewed in June (i.e., at the end of the season). The main message of the table is that FCB titles have a strong effect (only on men) increasing regional identity in Catalonia (+12%) and boosting preferences for higher autonomy (+5%) and support for regional self-determination (+12%).

---

[10]Note that the CIS barometer is carried out during the first two weeks of each month and stops in August, hence limiting my ability to study the persistence of the effects.

Figure 2: Impact of FCB titles (that season) on Catalan's identity and political preferences

Figure 2.a: Identity
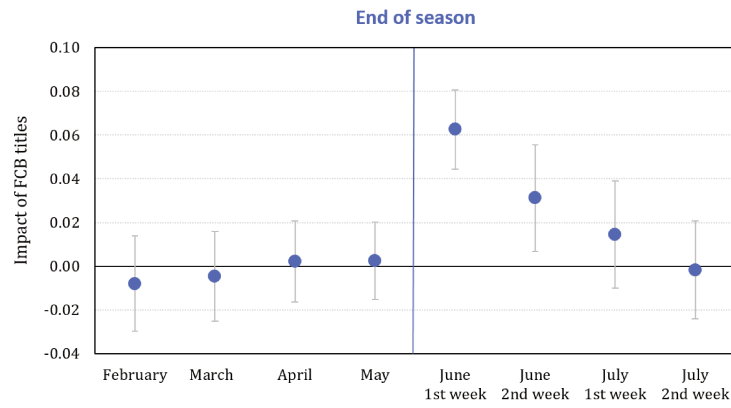
Figure 2.b: Favors higher autonomy
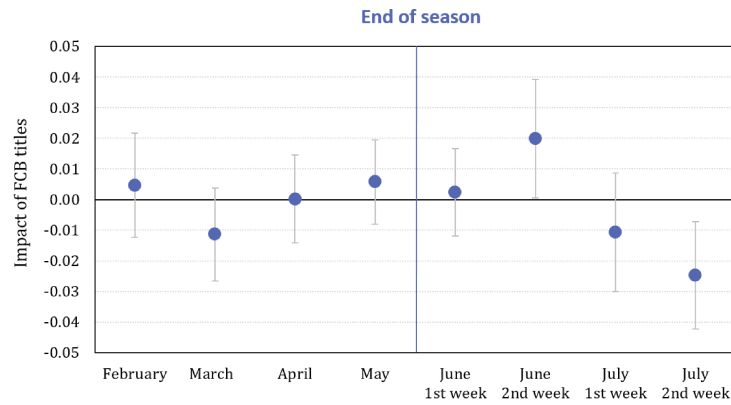
Figure 2.c: Favors the right of self-determination
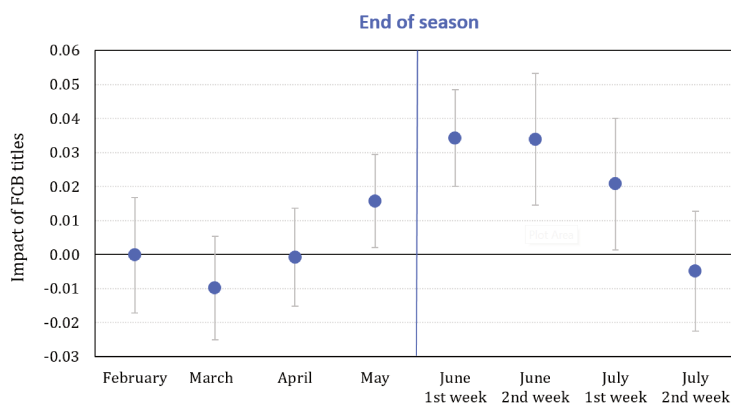
Note: This figure plots the impact of FCB titles on identity and political preferences of Catalan men over time (i.e., the results of estimating equation (2) for different time intervals). The outcomes are dummies that represent identifying more with the region than with Spain (Panel A), favoring higher regional autonomy (Panel B) and favoring the right of self-determination for regions (Panel C). As an illustration, Panel C shows that one additional FCB title increases support for regional self-determination among Catalan men by 3.5 percentage points right after the season ends in June. Dotted lines represent 95% confidence intervals based on standard errors clustered at the *region × year* level.

190

Table 6: Impact of FCB titles on identity and political preferences

	1	2	3	4	5	6	7
	Panel A: Strong Catalan identity						
Catalan x FCB titles	0.029	0.028	0.029	0.028	0.052	0.005	0.059
	0.013	0.005	0.005	0.005	0.006	0.008	0.014
Observations	22,369	22,369	22,369	22,369	10,834	11,535	2,983
R2	0.066	0.155	0.156	0.165	0.173	0.160	0.097
	Panel B: Favors higher autonomy						
Catalan x FCB titles	0.014	0.013	0.013	0.009	0.032	-0.014	0.030
	0.012	0.004	0.004	0.005	0.009	0.008	0.013
Observations	18,006	18,006	18,006	18,006	9,086	8,920	2,581
R2	0.140	0.198	0.199	0.242	0.255	0.232	0.151
	Panel C: Favors right of self-determination						
Catalan x FCB titles	0.009	0.009	0.009	0.007	0.044	-0.030	0.037
	0.010	0.007	0.008	0.007	0.005	0.013	(0.009
Observations	18,006	18,006	18,006	18,006	9,086	8,920	2,581
R2	0.131	0.173	0.173	0.200	0.214	0.188	0.151
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Regional trends	No	No	Yes	No	No	No	No
Individual controls	No	No	No	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Men	Women	Men
					[All regions]	[All regions]	[Identity regions]

Note: This table shows the impact of FCB titles on identity and political preferences (i.e., the results from estimating equation 1). The outcomes are dummies that represent identifying more with the region than with Spain (Panel A), favoring higher regional autonomy (Panel B) and favoring the right of self-determination for regions (Panel C). Each combination of Panel and column represents a separate regression. Catalan is a dummy for being in Catalonia and FCB titles is the number of major titles recently won by FCB when the individual was interviewed in June. As an example, column (1) in Panel A shows that each additional FCB titles increases Catalan identity by 2.9 percentage points. Controls include age and dummies for being male, married, catholic, college education and having left orientation. The standard errors are clustered at the *region* $\times$ *year* level. *, **, *** indicate p-values below 10%, 5%, an 1% respectively.

Columns (1) to (4) introduce different sets of controls and fixed effects. The baseline specification in column (4) includes year and regional fixed effects, and individual controls. As shown in column (4) of Panel A, each additional FCB title increases Catalan identification by around 2.8 percentage points at the end of the season. This effect seems mostly driven by men as displayed in column (5). On average, regional identity among Catalan men increases by 5.2 percentage points for every title FCB wins that season (i.e., an 11.6% rise relative to baseline). This estimate is similar in magnitude to the effect reported by (Depetris-Chauvin et al., 2020). They find that regional identification falls by 5.3 percentage points after a victory of the national team in African football. There is a concern that FCB titles could have coincided with years in which regional identity was on the rise in Spain. To better account for this, column (7) relies on regions with a salient identity; Catalonia, Basque Country, and Galicia. The results are unchanged whether one compares Catalan to the rest of Spanish or to specific regions with strong identity feelings.

Analogously to the findings on the impact of FCB matches, Panels B and C show that FCB titles not only affect Catalan identity but also have an impact on political preferences. Column (5) in Panel B shows that the share of men favoring higher autonomy increases by 3.2 percentage points for every FCB title (i.e., a 5.1% rise). Similarly, Column (5) in Panel C shows that the share of men favoring the right of self-determination increases by 4.4. percentage points for every FCB title (i.e., a 11.7% rise). As before, restricting the sample to regions with strong identity feelings barely affects the results.

ROBUSTNESS: To further assess the robustness of my empirical approach and findings, I perform a number of (i) placebo, (ii) sensitivity, and (iii) falsification tests. The sensitivity analysis consists in excluding years and regions sequentially. As shown in Figure A3, the estimated coefficients are not influenced by any year in particular. Similarly, Figure A4 shows that the results are not driven by specific regions, as the estimates remain virtually unchanged when excluding each of them.

Next, I repeat the analysis using placebo variables instead of the number of FCB titles. The results are summarized in Table A1. In columns (1) and (2) I consider the titles obtained in the next or precedent season. None of them seem to have an impact on idenitity or preferences. In columns (3) and (4) I examine the influence of less important titles (e.g., Supercups, FIFA Club World Cup) and the performance of the main rival team, Real Madrid. Less relevant titles or those won by

Real Madrid do not seem to have an effect on Catalan's sentiments. Similarly, in columns (5) and (6) I look at the performance of the main teams in other regions with a strong identity, such as Celta de Vigo in Galicia and Athletic Club in the Basque Country. Given that they did not win any trophy in the period considered, I rely on their final position in the league. As expected, their performance is not associated with changes in Catalan identity or preferences. I also address the concern that FCB may be more successful when there are more Catalan players in the main squad, and these, rather than the titles could be driving the effects. To account for this, for each season I count the number of Catalan born players having participated in most of the matches (this variable ranges from 3 to 7). As shown in column (7), the number of Catalan players in the main squad does not seem to influence identity or preferences.

Finally, I carry out a falsification exercise by excluding Catalan from the data and repeating the analysis focusing on each region one by one (i.e., the thought of experiment of seeing people from region r as being Catalan). The results are summarized in Table A2. The estimates of the interaction term are very closed to zero and insignificant in almost all regressions. For comparison purposes, I include the results for Catalonia in column (9). No other region presents a positive and significant effect of FCB titles on regional identity as Catalan do. The sole exception is La Rioja in column (17), yet the large and positive effect on identity seems at odds with the large and negative coefficient for the outcomes regarding preferences for autonomy and self-determination.

5.3   Additional results: FCB titles, euphoria and trust in politicians

This section explores some of the mechanisms that could explain why FCB titles affect identity and political views among Catalan men, focusing on dimensions such as optimism and trust.

As discussed by Healy et al. (2010), sports could trigger emotional responses that in turn affect behavior. They provide evidence pointing to changes in euphoria after upset losses. To examine this hypothesis, I exploit additional information from the CIS barometer on political and economic prospects as well as life satisfaction. Panel A of Table 3 summarizes the results of estimating equation (1) for these outcomes. Column (1) shows that FCB titles do not seem to influence subjective well-being (i.e., self-reported life satisfaction in a 1-10 scale). Yet, the information on life satisfaction was not collected in three years in the period of study, posing a limit in the interpretation of this result. As shown in columns (2) and (5), FCB titles increase optimism reflected by a more positive evaluation of the current political and economic situation of the country. Each additional title increases the share of Catalan who think the political and economic situation is good by around 4 percentage points. Similarly, the political and economic prospects also improve. FCB titles reduce the share of Catalan who think that next year the political (economic) situation will be worse than today by 3.5 (1.9) percentage points. Depetris-Chauvin et al. (2020) also examined similar outcomes and did not find any evidence that victories of African national teams increased subjective well-being or economic prospects (neither personal nor for the nation).

Another related channel that could mediate the effect of FCB titles on political preferences is by affecting trust in others or views on politicians (Depetris-Chauvin et al., 2020). The CIS barometer gathers information on trust in the President of the Central Government and the leader of the opposition. Overall, trust in the President is very low, with 33% of individuals reporting low trust, and 49% no trust at all. The results, summarized in Panel B, show that FCB titles reduce mistrust in the president of the Government while having the opposite impact on the leader of the opposition. On average, each additional title reduces the share of Catalan with no trust at all in the President by 6.3 percentage points, while increases the share of individuals with no trust in the opposition leader by 5.5 percentage points. The finding that FCB titles foster trust in the Central Government may seem surprising at first sight. Yet, in the context of the Catalan-Spanish conflict, the Catalan independence may only be achieved with the collaboration of the Spanish state, and

thus greater hopes for independence may come along with higher trust in the Government. As before, Depetris-Chauvin et al. (2020) do not find any evidence that football matches in Africa increased trust in the Government, however, they do find robust evidence on increase inter-ethnic trust.

Table 7: Impact of FCB titles on subejctive well-being, prospects and trust in politicians

	Panel A: Life satisfaction, political and economic prospects						
	Well-being (1-10)	Pol. sit. good	Pol. sit. t-1 worse	Pol. sit. t+1 worse	Econ. sit. good	Econ. sit. t-1 worse	Econ. sit. t+1 worse
	1	2	3	4	5	6	7
Catalan x FCB titles	-0.016	0.041	-0.044	-0.035	0.04	0.001	-0.019
	0.011	0.019	0.010	0.010	0.028	0.016	0.009
Observations	8,404	13,368	13,751	13,751	13,510	13,572	13,572
R2	0.09	0.065	0.029	0.022	0.162	0.136	0.04
	Panel B: Trust in the President of Central Government and Oposition leader						
		Some trust in Gov.	Low trust in Gov.	No trust in Gov.	Some trust in Gov.	Low trust in Gov.	No trust in Gov.
		2	3	4	5	6	7
Catalan x FCB titles		0.02	0.033	-0.063	-0.021	-0.033	0.055
		0.017	0.015	0.029	0.009	0.025	0.032
Observations		12,341	12,341	12,341	11,143	11,143	11,143
R2		0.03	0.011	0.058	0.022	0.011	0.035
Year and region fixed effecs		Yes	Yes	Yes	Yes	Yes	Yes
Individual controls		Yes	Yes	Yes	Yes	Yes	Yes

Note: This table shows the impact of FCB titles on economic and the political prospects (Panel A) and trust in politicians (Panel B) among Catalan men (i.e., the results from estimating equation 1). Each combination of Panel and column represents a separate regression. Catalan is a dummy for being in Catalonia and FCB titles is the number of major titles recently won by FCB when the individual was interviewed in June. The wording t-1 and t+1 stand for the previous (next) year (e.g.," Pol. Sit. t-1" means the respondent thinks the political situation a year ago was worse than today). As an example, column (7) in Panel A shows that each additional FCB titles decreases the share of Catalan who think that the economic situation next year will be worse than today by 1.9 percentage points. The standard errors are clustered at the *region × year* level. *, **, *** indicate p-values below 10%, 5%, an 1% respectively.

6 Impact of FCB matches on electoral results [Very preliminary]

6.1 Empirical Strategy

To estimate the impact of FCB matches on electoral outcomes I exploit a panel of all Catalan municipalities between 1979 and 2019. Since all elections take place on Sunday, I define *treated* elections as those having a close match (i.e., the day before). I introduce a measure of FCB support by relying on the number of fan clubs (*peñas*) across Catalan districts (*comarcas*). This brings spatial variation in the intensity of the shock, with the areas having more fan clubs being a priori more likely to be affected by FCB matches. Combining the information on close matches with the measure of FCB support allows me to include election fixed effects that capture the effect of yearly-shocks common to the whole region. The estimating equation is:

$$V_{mde} = \theta_m + \theta_e + \phi \, FCBFans_d \times Closematch'_e + v_{mde} \,, \tag{3}$$

where m, d and e index municipalities, districts and elections, θ_m and θ_e are municipality and election fixed effects, $FCBFans_m$ represents the support for FCB in district d (i.e., the number of FCB fan clubs), $Closematch'_e$ is a vector of three dummy variables representing the outcome of the match played the day before election e (e.g., win_e, $draw_e$, and $loss_e$). I cluster the standard errors v_{mde} at the municipality level. The coefficient ϕ represents how the electoral result V_{mde} changes if there was a close match, depending on its outcome and the extent of FCB support. The *comparison* group is therefore elections without a match the day before, this includes elections with a match on that same day (Sunday), or with the nearest match being several days before. I will provide the results with alternative definitions of the *comparison* group.

To interpret the estimates of equation (3) as causal I rely on the assumption that the of matches and elections are unrelated. This is quite plausible given that the calendar for the whole football season is set in August and whether FCB will play a Saturday or a Sunday in a given weekend is as good as random. Similarly elections are often scheduled without knowing when (or where) FCB will play.

6.2 Results: FCB matches and electoral results

Table 8 summarizes the results about the impact of FCB matches on electoral results. Panel focuses on electoral participation while Panel B on the vote-share for parties with a stronger identity component. The specifications in column 1 compare elections with a match the day before with all the rest. In column 2 elections with matches on that same day are excluded. In column 3 the sample comprises elections with a match the day before and on the very same same day. Finally, column 4 includes only election with a match the day before.

As shown in Panel A, regardless of the comparison group, FCB matches seem to affect electoral participation. For every FCB fan club in the district, participation falls by 0.2 percentage points if FCB did not win the day before. Note that the median number of FCB fan clubs is 10, which would imply a fall in participation of 2 percentage points if FCB did not win. Panel B however shows a blurry picture about the potential impact of FCB matches on the support for particular parties. The point estimates change considerably depending on the elections considered in the comparison group. Besides a null effect of FCB matches, another possibility is that individuals shift their support across parties in a way that escapes the way in which parties were aggregated under the identity parties.

Table 8: Impact of FCB matches on electoral results

	Panel A: Participation			
	1	2	3	4
FCB win [day before] X Fan clubs	-0.0002	-0.0001	-0.0003	
	0.0001	0.0002	0.0002	
FCB draw [day before] X Fan clubs	-0.002***	-0.002***	-0.002***	
	0.0002	0.0002	0.0002	
FCB loss [day before] X Fan clubs	-0.002***	-0.002***	-0.002***	-0.002***
	0.0003	0.0003	0.0003	0.0003
Observations	17,959	11,610	11,447	8,354
R2	0.118	0.115	0.127	0.110
	Panel B: Vote share of identity parties			
	1	2	3	4
FCB win [day before] X Fan clubs	-0.0004	-0.002***	0.00003	
	0.0003	0.0004	0.0004	
FCB draw [day before] X Fan clubs	0.003***	0.002***	0.003***	
	0.0004	0.001	0.001	
FCB loss [day before] X Fan clubs	-0.0003	-0.001**	0.0001	-0.001
	0.001	0.001	0.001	0.001
Observations	17,959	11,610	11,447	8,354
R2	0.024	0.024	0.022	0.016
Municipality and election fixed effects	Yes	Yes	Yes	Yes
Province x year fixed effects	Yes	Yes	Yes	Yes
Comparison group	Elections without a match day before	Exclude elections with matches on that day	Elections with a match that day	Other elections with a match the day before

Note: This table summarizes the results on the impact of FCB matches on electoral results in Catalonia. Participation refers to the ratio of cast votes over electorate in the municipality. I group as Catalan identity parties the following: ERC, CDC, CiU, JxC, CUP. Treated elections are those in which FCB played the day before. FCB fan clubs refers to the number of *peñas* in the district (*comarca*). Standard errors clustered at the municipality level.*, **, *** indicate p-values below 10%, 5%, an 1% respectively.

7 Conclusion

This paper shows that when there is a strong link between a sports club and an identity trait, its performance can temporary shape identity, political preferences and ultimately electoral results. Leveraging survey data from various sources, I find that FCB performance has a strong effect on Catalan identification (in detriment of Spanish) and support for self-determination and Catalan secession. In particular, each additional FCB title is estimated to increase the share of individuals reporting a strong sense of Catalan identity by 11% and preferences over self-determination by 12%. Similarly, in the rare event of an FCB defeat, individuals surveyed one day after the match (rather than before) are 11% less likely to identify strongly with Catalonia and 15% less likely to support Catalan independence. FCB matches also affect electoral outcomes in areas with more supporters. In particular, FCB defeats one day before an election reduce electoral participation by 0.2 percentage points for every FCB fan club in the district (being the median around ten).

I also show that FCB performance seems to affect Catalan's euphoria, manifested by a more optimistic evaluation of the current economic and political situation as well as future prospects. In parallel, FCB triumphs reduce mistrust in the Central Government, which could indicate greater hopes for dialogue and progress regarding the Catalan independence process. The social environment also seems to play a role, as the impact of defeats becomes stronger in areas with more FCB fan clubs and a higher density of pubs per capita.

My findings have major policy implications. First, they show that by driving emotions, sports can temporarily influence identity feelings and change political preferences. Politicians may be tempted to use sports strategically for their own benefit or to impulse a given agenda. As a result, sports can serve to both unite and polarize a country depending on the ideals they portrait. This highlights the importance of separating as much as possible sports from politics. Even if the emotional trigger of sports is short-lived, my findings suggest that they can influence electoral results, which in turn, can alter the politicians elected or the policies implemented having long-lasting consequences. This is particularly relevant in the context of high-stake decisions, such as referendums, since the final result could be shaped by factors as irrelevant as a football match. While undoubtedly sports are a central part of people's life and can bring plenty of positive aspects, it is on us to be conscious of all its potential effects and channel them towards the best ends.

References

Alesina, A. and Zhuravskaya, E. (2011). Segregation and the quality of government in a cross section of countries. *American Economic Review*, 101(5):1872–1911.

Arnold, F. and Freier, R. (2016). Only conservatives are voting in the rain: Evidence from german local and state elections. *Electoral Studies*, 41:216–221.

Augenblick, N. and Nicholson, S. (2016). Ballot position, choice fatigue, and voter behaviour. *The Review of Economic Studies*, 83(2):460–480.

Bandiera, O., Mohnen, M., Rasul, I., and Viarengo, M. (2019). Nation-building through compulsory schooling during the age of mass migration. *The Economic Journal*, 129(617):62–109.

Banerjee, A., Iyer, L., and Somanathan, R. (2005). History, social divisions, and public goods in rural india. *Journal of the European Economic Association*, 3(2-3):639–647.

Bazzi, S., Gaduh, A., Rothenberg, A. D., and Wong, M. (2019). Unity in diversity? how intergroup contact can foster nation building. *American Economic Review*, 109(11):3978–4025.

Besley, T. and Reynal-Querol, M. (2014). The legacy of historical conflict: Evidence from africa. *American Political Science Review*, 108(2):319–336.

Blouin, A. and Mukand, S. W. (2019). Erasing ethnicity? propaganda, nation building, and identity in rwanda. *Journal of Political Economy*, 127(3):1008–1062.

Cantoni, D., Chen, Y., Yang, D. Y., Yuchtman, N., and Zhang, Y. J. (2017). Curriculum and ideology. *Journal of Political Economy*, page 55.

Card, D. and Dahl, G. B. (2011). Family violence and football: The effect of unexpected emotional cues on violent behavior*. *The Quarterly Journal of Economics*, 126(1):103–143.

Clots-Figueras, I. and Masella, P. (2013). Education, language and identity. *The Economic Journal*, 123(570):332–357.

Cristancho, C., Anduiza, E., and González, S. (2020). Space in contentious politics: Protest supply and individual level mobilisation. *mimeo*.

Dehdari, S. H. and Gehring, K. (2019). The origins of common identity: Evidence from alsace-lorraine.

Dell, M. and Querubin, P. (2018). Nation building through foreign intervention: Evidence from discontinuities in military strategies. *The Quarterly Journal of Economics*, 133(2):701–764.

Depetris-Chauvin, E., Durante, R., and Campante, F. (2020). Building nations through shared experiences: Evidence from african football. *American Economic Review*, 110(5):1572–1602.

Duncan, B., Mansour, H., and Rees, D. I. (2017). It's just a game: The super bowl and low birth weight. *Journal of Human Resources*, 52(4):946–978.

Eren, O. and Mocan, N. (2018). Emotional judges and unlucky juveniles. *American Economic Journal: Applied Economics*, 10(3):171–205.

Foer, F. (2004). *How soccer explains the world: An unlikely theory of globalization*. Harper Collins New York.

Fouka, V. (2020). Backlash: The unintended effects of language prohibition in u.s. schools after world war i. *The Review of Economic Studies*, 87(1):204–239.

Fujiwara, T., Meng, K., and Vogl, T. (2016). Habit formation in voting: Evidence from rainy elections. *American Economic Journal: Applied Economics*, 8(4):160–88.

Galeano, E. (2014). *Soccer in sun and shadow*. Open Road Media.

Healy, A. J., Malhotra, N., and Mo, C. H. (2010). Irrelevant events affect voters' evaluations of government performance. *Proceedings of the National Academy of Sciences*, 107(29):12804–12809.

Liberini, F., Redoano, M., and Proto, E. (2017). Happy voters. *Journal of Public Economics*, 146:41–57.

Lind, J. T. (2019). Spurious weather effects. *Journal of Regional Science*, 59(2):322–354.

Lindo, J. M., Swensen, I. D., and Waddell, G. R. (2012). Are big-time sports a threat to student achievement? *American Economic Journal: Applied Economics*, 4(4):254–74.

Meier, A. N., Schmid, L., and Stutzer, A. (2019). Rain, emotions and voting for the status quo. *European Economic Review*, 119:434–451.

Montolio, D. and Planells-Struse, S. (2016). How time shapes crime: The temporal impacts of football matches on crime. *Regional Science and Urban Economics*, 61:99–113.

Mousa, S. (2020). Building social cohesion between christians and muslims through soccer in post-isis iraq. *Science*, 369(6505):866–870.

Munyo, I. and Rossi, M. A. (2013). Frustration, euphoria, and violent crime. *Journal of Economic Behavior & Organization*, 89:136–142.

Passarelli, F. and Tabellini, G. (2017). Emotions and political unrest. *Journal of Political Economy*, 125(3):903–946.

Ramonet, I. (1998). Un fait social total. « Le football, fait social total », par Ignacio Ramonet (accès libre, Manière de voir, mai 1998) // https://www.monde-diplomatique.fr/55358.

Rodon, T. and Guinjoan, M. (2018). When the context matters: Identity, secession and the spatial dimension in catalonia. *Political Geography*, 63:75–87.

Shobe, H. (2008a). Football and the politics of place: Football club barcelona and catalonia, 1975–2005. *Journal of cultural geography*, 25(1):87–105.

Shobe, H. (2008b). Place, identity and football: Catalonia, catalanisme and football club barcelona, 1899–1975. *National identities*, 10(3):329–343.

Tomás, M. and Porta, F. (2016). *Barça inédito: 800 historias de la Historia*. Córner.

Úbeda-Colomer, J., Molina, P., and Villamón, M. (2017). La construcción de identidades colectivas a través del fútbol: cuando los nacionalismos juegan en los estadios. *Materiales para la Historia del Deporte*, (15):45–56.

Van Winden, F. (2015). Political economy with affect: On the role of emotions and relationships in political economics. *European Journal of Political Economy*, 40:298–311.

Voigtländer, N. and Voth, H.-J. (2015). Nazi indoctrination and anti-semitic beliefs in germany. *Proceedings of the National Academy of Sciences*, 112(26):7931–7936.

Vázquez-Montalbán, M. (1997). Le football, religion laïque en quête d'un nouveau dieu. « Le football, religion laïque en quête d'un nouveau Dieu », par Manuel Vazquez Montalban (accès libre, août 1997) // https://www.monde-diplomatique.fr/4899.

Online Appendix

A   Additional Figures and Tables

Figure 1: Distribution of identity across Spanish regions
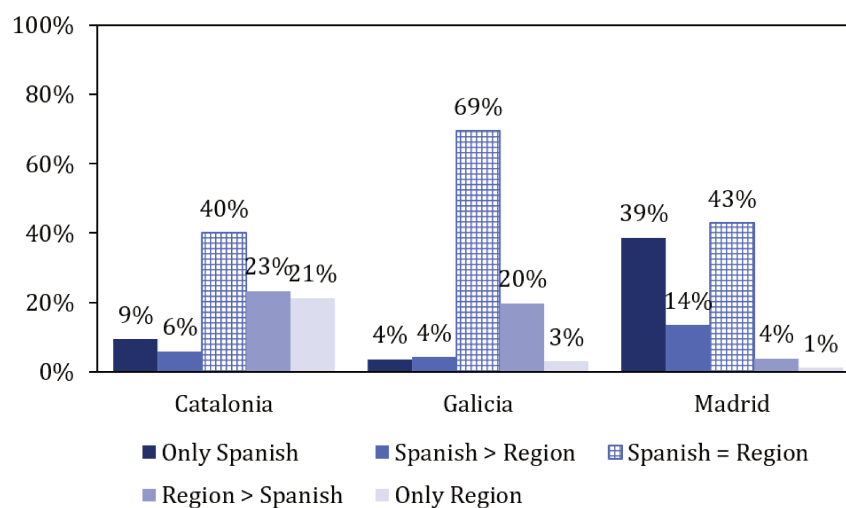
Note: Own elaboration based on CIS barometers, 2012-2018. Return to page 10.

Figure 2: Distribution of preferences on decentralization across Spanish regions
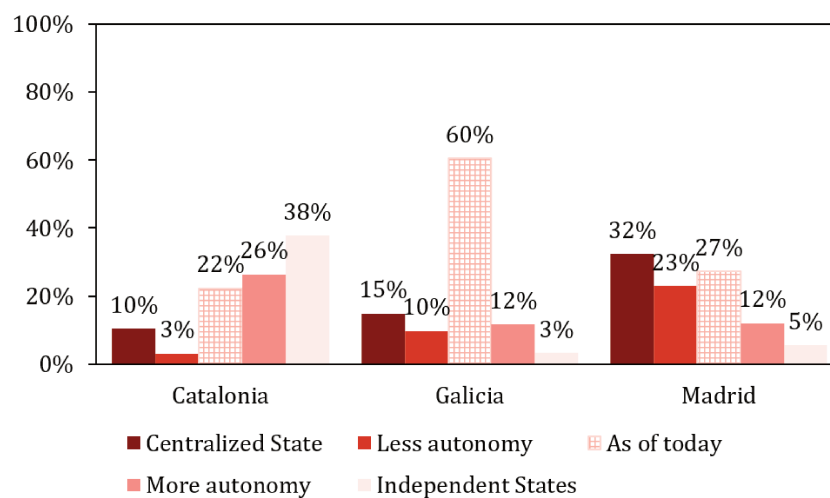
Note: Own elaboration based on CIS barometers, 2012-2018. Return to page 10.

Figure 3: FCB fan clubs across Catalan municipalities (*peñas*)
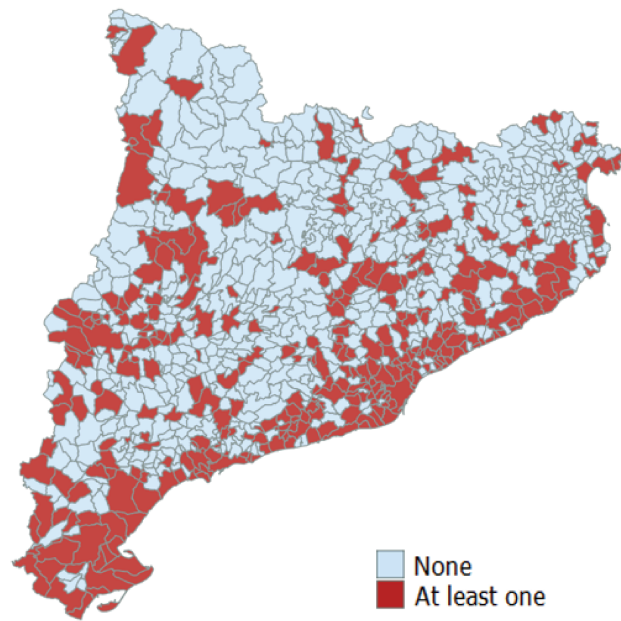
Note: This map plots the municipalities with at least one FCB *peña*. Information retrieved from FCB official sources. Return to page 27.

Figure 4: Sensitivity excluding years: Impact of FCB titles on identity and preferences

Figure 5: Sensitivity excluding regions: Impact of FCB titles on identity and preferences

Table 1: Impact of FCB titles on interview characteristics

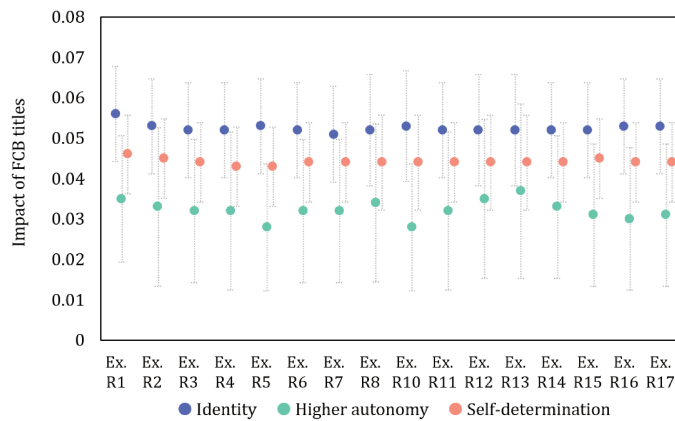|	Rush	Complaints	Good talk	Sincerity	Length (min)
	1	2	3	4	5
Catalan x FCB titles	-0.003	-0.001	-0.029***	-0.015	-0.390***
	0.006	0.002	0.013	0.010	0.181
Observations	14,038	14,038	14,038	14,038	12,500
R2	0.018	0.021	0.048	0.043	0.467
Year and region fixed effecs	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes

Note: This table summarizes the impact of FCB titles on various characteristics of the interview. Rush represents whether the respondent urged the interviewer to go faster, Complaints whether the respondent complained about any particular question, Good talk the assessment of the conversation, Sincerity the assessment over the sincerity of the respondent, and Length the duration of the interview in minutes. Individual controls include age and dummies for being married, employed, Catholic, left-orientation and having college education. The standard errors are clustered at the *region × year* level.*, **, *** indicate p-values below 10%, 5%, an 1% respectively.

Table 2: Robustness checks: Impact of FCB titles on identity and preferences

	Panel A: Identity						
	1	2	3	4	5	6	7
Catalan x Interaction	0.001	-0.01	-0.009	0.011	-0.002	0.001	-0.008
	0.008	0.006	0.008	0.01	0.001	0.002	0.005
Observations	10,834	10,834	10,834	10,834	10,834	10,834	10,834
R2	0.173	0.173	0.173	0.173	0.173	0.173	0.173
	Panel B: Favors higher autonomy						
	1	2	3	4	5	6	7
Catalan x Interaction	0.017	-0.009	0.008	-0.006	0.001	0.0003	-0.004
	0.006	0.006	0.006	0.008	0.002	0.002	0.004
Observations	9,086	9,086	9,086	9,086	9,086	9,086	9,086
R2	0.256	0.255	0.255	0.255	0.255	0.255	0.255
	Panel C: Favors right of self-determination						
	1	2	3	4	5	6	7
Catalan x Interaction	0.008	0.008	-0.005	0.013	-0.004	-0.003	-0.006
	0.007	0.005	0.009	0.008	0.001	0.001	0.005
Observations	9,086	9,086	9,086	9,086	9,086	9,086	9,086
R2	0.214	0.214	0.214	0.214	0.214	0.214	0.214
Interaction	FCB titles t+1	FCB titles t-1	FCB irrelevant titles	Real Madrid titles	Celta Liga rank	Athletic Liga rank	FCB Catalan players
Year and region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Return to page .

Table 3: Falsification test: Impact of FCB titles on other regions' identity and preferences

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Panel A: Identity																	
Region	0.002	-0.054	-0.019	0.146	0.229	-0.041	-0.084	-0.101	0.201	-0.063	-0.025	0.109	-0.16	-0.124	0.308	0.396	-0.139
	0.043	0.046	0.034	0.09	0.06	0.04	0.026	0.030	0.027	0.029	0.047	0.047	0.032	0.033	0.063	0.06	0.072
Region x FCB titles	0.023	0.001	-0.014	-0.015	0.013	-0.007	-0.018	-0.01	0.052	0.001	-0.013	-0.02	0.001	-0.006	-0.043	0.002	0.069
	0.019	0.019	0.022	0.061	0.027	0.022	0.014	0.016	0.014	0.017	0.023	0.031	0.014	0.014	0.023	0.025	0.045
Observations	9,211	9,211	9,211	9,211	9,211	9,211	9,211	9,211	10,834	9,211	9,211	9,211	9,211	9,211	9,211	9,211	9,211
R2	0.021	0.019	0.019	0.021	0.042	0.019	0.024	0.026	0.091	0.022	0.019	0.022	0.038	0.023	0.026	0.075	0.019
Panel B: Favors higher autonomy																	
Region	-0.001	-0.038	-0.032	0.153	0.283	0.026	-0.08	-0.088	0.362	0.033	-0.097	-0.096	-0.155	-0.144	0.244	0.407	0.031
	0.035	0.032	0.031	0.112	0.036	0.08	0.034	0.023	0.024	0.032	0.063	0.034	0.035	0.064	0.077	0.041	0.104
Region x FCB titles	0.014	0.024	-0.01	-0.012	-0.079	0.004	0.003	0.021	0.032	-0.034	0.006	0.036	0.025	0.037	-0.046	-0.037	-0.1
	0.016	0.023	0.023	0.059	0.029	0.045	0.017	0.016	0.013	0.018	0.042	0.014	0.02	0.035	0.04	0.019	0.056
Observations	7,631	7,631	7,631	7,631	7,631	7,631	7,631	7,631	9,086	7,631	7,631	7,631	7,631	7,631	7,631	7,631	7,631
R2	0.086	0.086	0.086	0.089	0.098	0.086	0.088	0.088	0.213	0.087	0.088	0.088	0.097	0.088	0.09	0.126	0.087
Panel C: Favors the right of self-determination																	
Region	0.022	-0.014	-0.022	0.134	0.052	0.062	-0.041	-0.039	0.225	-0.014	-0.061	-0.043	-0.076	-0.049	0.105	0.258	0.032
	0.021	0.02	0.014	0.092	0.017	0.037	0.022	0.014	0.017	0.019	0.019	0.021	0.021	0.02	0.065	0.031	0.052
Region x FCB titles	0.01	0.026	0.003	-0.033	-0.025	-0.026	0.01	0.002	0.044	0.0004	-0.0005	0.005	-0.005	-0.011	0.045	-0.006	-0.07
	0.01	0.017	0.009	0.047	0.012	0.024	0.013	0.008	0.009	0.011	0.01	0.012	0.01	0.011	0.028	0.014	0.029
Observations	7,631	7,631	7,631	7,631	7,631	7,631	7,631	7,631	9,086	7,631	7,631	7,631	7,631	7,631	7,631	7,631	7,631
R2	0.07	0.068	0.068	0.072	0.069	0.068	0.069	0.069	0.185	0.068	0.07	0.069	0.081	0.07	0.076	0.117	0.069
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Region definitions: R1 Andalucía, R2 Aragón, R3 Asturias, R4 Balearic Islands, R5 Canary Islands, R6 Cantabria, R7 Castilla y León, R8 Castilla-La Mancha, R9 Catalonia, R10 Valencia, R11 Extremadura, R12 Galicia, R13 Madrid, R14 Murcia, R15 Navarra, R16 Basque Country, R17 Rioja.

B Description of FCB matches

Table 1: Summary of FCB matches in the final sample

Match date	Rival	Competition	FCB goals	Rival's goals	Defeat	Match date	Rival	Competition	FCB goals	Rival's goals	Defeat
3/23/2014	Real Madrid	Liga	4	3	0	2/7/2016	Levante UD	Liga	2	0	0
3/26/2014	Celta Vigo	Liga	3	0	0	2/23/2016	Arsenal FC	Champions	2	0	0
3/29/2014	Espanyol	Liga	1	0	0	2/28/2016	Sevilla FC	Liga	2	1	0
4/1/2014	Atlético Madrid	Champions	1	1	0	10/2/2016	Celta Vigo	Liga	3	4	1
4/5/2014	Real Betis	Liga	3	1	0	10/19/2016	Manchester City	Champions	4	0	0
4/9/2014	Atlético Madrid	Champions	0	1	1	10/22/2016	Valencia CF	Liga	3	2	0
4/12/2014	Granada CF	Liga	0	1	1	10/29/2016	Granada CF	Liga	1	0	0
4/16/2014	Real Madrid	Copa	1	2	1	11/1/2016	Manchester City	Champions	1	3	1
11/1/2014	Celta Vigo	Liga	0	1	1	12/18/2016	Espanyol	Liga	4	1	0
11/5/2014	AFC Ajax	Champions	2	0	0	3/8/2017	Paris Saint-Germain	Champions	6	1	0
11/8/2014	UD Almería	Liga	2	1	0	3/12/2017	Deportivo La Coruña	Liga	1	2	1
11/22/2014	Sevilla FC	Liga	5	1	0	3/19/2017	Valencia CF	Liga	4	2	0
12/10/2014	Paris Saint-Germain	Champions	3	1	0	10/18/2017	Olympiakos Piräus	Champions	3	1	0
12/13/2014	Getafe CF	Liga	0	0	0	10/21/2017	Málaga CF	Liga	2	0	0
2/8/2015	Athletic Bilbao	Liga	5	2	0	10/24/2017	Real Murcia	Copa	3	0	0
2/11/2015	Villarreal CF	Copa	3	1	0	10/28/2017	Athletic Bilbao	Liga	2	0	0
2/15/2015	Levante UD	Liga	5	0	0	1/11/2018	Celta Vigo	Copa	5	0	0
2/21/2015	Málaga CF	Liga	0	1	1	1/14/2018	Real Sociedad	Liga	4	2	0
2/24/2015	Manchester City	Champions	2	1	0	1/17/2018	Espanyol	Copa	0	1	1
2/28/2015	Granada CF	Liga	3	1	0	1/21/2018	Real Betis	Liga	5	0	0
6/6/2015	Juventus	Champions	3	1	0	1/25/2018	Espanyol	Copa	2	0	0
10/17/2015	Rayo Vallecano	Liga	5	2	0	1/28/2018	CD Alavés	Liga	2	1	0
10/20/2015	BATE Borisov	Champions	2	0	0	4/7/2018	CD Leganés	Liga	3	1	0
10/25/2015	SD Eibar	Liga	3	1	0	4/10/2018	AS Roma	Champions	0	3	1
10/28/2015	CF Villanovense	Copa	0	0	0	4/14/2018	Valencia CF	Liga	2	1	0
11/21/2015	Real Madrid	Liga	4	0	0	4/17/2018	Celta Vigo	Liga	2	2	0
11/24/2015	AS Roma	Champions	6	1	0	4/21/2018	Sevilla FC	Copa	5	0	0
2/3/2016	Valencia CF	Copa	7	0	0						

Note: This table summarizes all FCB matches in the final sample (i.e., all matches having individuals interviewed in the CEO barometer the day before and after the match).

Table 2: Summary of all elections in Catalonia and the previous FCB match

Election type	Election date	Close match	Last match date	Rival	FCB goals	Rival's goals	FCB defeat
Congress	6/15/1977	0	5/22/1977	Real Betis	3	1	0
Congress	3/1/1979	0	2/23/1979	Hercules FC	3	0	0
Municipal	4/3/1979	0	3/25/1979	Burgos CF	0	1	1
Regional	3/20/1980	1	3/19/1980	Valencia FC	3	4	1
Congress	10/30/1982	0	10/24/1982	Sporting Gijon	1	1	0
Municipal	5/8/1983	0	5/6/1983	Real Sociedad	2	0	0
Regional	4/29/1984	0	4/29/1984	Atletico Madrid	2	1	0
Congress	6/22/1986	0	6/13/1986	Real Betis	2	0	0
Municipal	6/10/1987	0	6/7/1987	Espanyol Barcelona	2	1	0
Europeans	6/10/1987	0	6/7/1987	Espanyol Barcelona	2	1	0
Regional	5/29/1988	0	5/22/1988	CA Osasuna	1	1	0
Europeans	6/15/1989	0	6/10/1989	Atletico Madrid	3	0	0
Congress	10/29/1989	1	10/28/1989	CD Tenerife	3	0	0
Municipal	5/26/1991	0	5/26/1991	CD Logrones	2	0	0
Regional	3/15/1992	1	3/14/1992	Atletico Madrid	2	2	0
Congress	6/6/1993	1	6/5/1993	Sevilla FC	2	1	0
Europeans	6/12/1994	0	5/18/1994	AC Milan	0	4	1
Municipal	5/28/1995	1	5/27/1995	Real Madrid	1	0	0
Regional	11/19/1995	1	11/18/1995	Albacete	3	0	0
Congress	3/3/1996	1	3/2/1996	Athletic Bilbao	0	0	0
Europeans	6/13/1999	0	6/13/1999	Real Betis	4	1	0
Municipal	6/13/1999	0	6/13/1999	Real Betis	4	1	0
Regional	10/17/1999	1	10/16/1999	CD Numancia	3	3	0
Congress	3/12/2000	1	3/11/2000	Athletic Bilbao	4	0	0
Municipal	5/25/2003	0	5/25/2003	Recreativo Huelva	3	1	0
Regional	11/16/2003	0	11/9/2003	Real Betis	2	1	0
Congress	3/14/2004	0	3/14/2004	Real Murcia	2	0	0
Europeans	6/13/2004	0	5/23/2004	Real Zaragoza	1	2	1
Regional	11/1/2006	1	10/31/2006	Chelsea FC	2	2	0
Municipal	5/27/2007	1	5/26/2007	Getafe CF	1	0	0
Congress	3/9/2008	0	3/4/2008	Celtic FC	1	0	0
Europeans	6/7/2009	0	5/30/2009	Deportivo La Coruna	1	1	0
Regional	11/28/2010	0	11/24/2010	Panathinaikos	3	0	0
Municipal	5/22/2011	1	5/21/2011	Malaga	3	1	0
Congress	11/20/2011	1	11/19/2011	Zaragoza	4	0	0
Regional	11/25/2012	0	11/25/2012	Levante UD	4	0	0
Europeans	5/25/2014	0	5/17/2014	Atletico Madrid	1	1	0
Municipal	5/24/2015	1	5/23/2015	Deportivo La Coruna	2	2	0
Regional	9/27/2015	1	9/26/2015	UD Las Palmas	2	1	0
Congress	12/20/2015	0	12/20/2015	River Plate	3	0	0
Congress	6/26/2016	0	5/22/2016	Sevilla FC	2	0	0
Regional	12/21/2017	0	12/17/2017	Deportivo La Coruna	4	0	0
Congress	4/28/2019	1	4/27/2019	Levante UD	1	0	0
Municipal	5/26/2019	1	5/25/2019	Valencia CF	1	2	1
Europeans	5/26/2019	1	5/25/2019	Valencia CF	1	2	1
Congress	11/10/2019	1	11/9/2019	Celta de Vigo	4	1	0

Note: This table summarizes all elections in Catalonia and the FCB match played the closest to the election game (always in retrospect, as some of these elections had a FCB on the election day).

Chapter 3

The Long-Term Impact of School Privatization on Well-Being:
Evidence from the 1981 Reform in Chile

The Long-Term Effects of School Privatization on Well-Being: Evidence from the 1981 Voucher Reform in Chile[*]

Abstract

In 1981 the Chilean Military Government introduced a new system of school choice that led to an exodus of middle-class students from public to subsidized private schools. Using historical administrative data of enrollment before and after the reform, I examine the long-term impact of this drastic rise in private enrollment on adult subjective well-being. The results show that children from disadvantaged families were negatively affected by the policy change. On average, being exposed to a rise in subsidized private enrollment of 10 percentage points during childhood is associated with a drop in adult life satisfaction of 6 percent of a standard deviation (e.g., equivalent to a fall in household income by 30 percent). Individuals from more advantaged family backgrounds experienced a modest increase in adult subjective well-being. The voucher reform did not seem to have a long-term impact on a variety of other outcomes such as health, schooling or wages. The mobility patterns of individuals (i.e., whether they left their municipality of birth) are however crucial to explain the results, suggesting that the change in peers and close contacts induced by the reform could have had persistent consequences on well-being.

Keywords: school segregation, subjective well-being, school choice, vouchers, Chile.

JEL Classification: I22, I24, I28, I31, Z13.

1 Introduction

In 1981, the Chilean Military Government introduced a nation-wide voucher program that profoundly change the school choice system. After this reform, parents could send their children to any public or subsidized private school without having to pay any fees. Schools would receive a flat-per-student subsidy depending on total enrollment and attendance. In just five years, the share of students attending subsidized private schools soared from 15 to 30 percent. This exodus was however characterized by a composition effect; subsidized private schools attracted mainly middle-class students, leading to a sharp increase in school segregation (Hsieh and Urquiola, 2006). While the main features of this school choice system have remained unchanged for more than three decades. Today, Chile presents one of the highest rates of private school enrollment and is one of the countries with greatest school segregation in the world (see Figures A.1 and A.2 in the Appendix). Despite the major implications of the voucher reform, to date there has been no assessment of its potential consequences in the long run.

This paper presents the first empirical evidence on the long-term effects of school vouchers on subjective well-being by studying the 1981 voucher reform in Chile. I combine detailed survey data with historical records of enrollment across municipalities and exploit the unequal penetration of private schools after 1981.[1] The identification strategy relies on the comparison of individuals born in the same municipality but differently exposed to the reform depending on their age at the time it was implemented. The results show that the voucher reform had a long-lasting impact on subjective well-being, harming individuals with low-educated parents while modestly benefiting those with highly educated parents. For individuals with low educated parents (i.e., bottom 40% in schooling), being exposed to a 10 percentage point increase in subsidized private enrollment at ages 6-14 led to a fall in adult subjective-well-being of around 6 percent of a standard deviation. This effect is equivalent to the impact of decreasing household income by 30 percent or schooling by 3 years. By contrast, for individuals with highly educated parents (i.e., top 20% in schooling), being exposed to a 10 percentage point increased in subsidized private enrollment is associated with an increase in adult subjective-well-being of around 3.8 percent of a standard deviation.

---

[1]Municipalities refer to Chilean *comunas*, the smallest administrative entity in the country. *Comunas* resemble US counties and vary significantly in size and population. As of 2012, the 10 percent least populous *comunas* had an average population of 1,900 people while the 10 percent most populous had an average population of 253,000 people. Chile is divided in 15 regions, 54 provinces and 346 *comunas*.

The impact of the reform differs remarkably depending on the mobility patterns of individuals. Among those with low-educated parents, the voucher reform had a long-term negative impact if they live in municipalities with a large entry of private schools irrespective of whether they were born or moved there. On average, a 10 percentage point increase in subsidized private enrollment is associated with a fall in adult life satisfaction of 9.3 to 11.1 percent of a standard deviation. Interestingly, for those who have moved to other municipalities the degree of penetration of private schools in their birth municipalities did not seem to have had any long-lasting effect.

This paper contributes to different strands of the literature. First, it is part of a large body of research assessing the effects of school choice and segregation on a variety of outcomes. Previous work has found that flexible school choice often leads to increases in segregation and exacerbates inequalities between children of different social backgrounds (Sorderstrom and Uusitalo, 2005: Bohlmark et al., 2016). My results point in this direction showing that the long-term impact of the vouchers on well-being was completely different depending on parental education. While all children had equal access to vouchers, those with lower-educated parents were much more likely to attend public schools and experience a drop in the average socio-economic status of their peers. Three decades later, these children display lower subjective well-being levels than other individuals with similar family background who were not exposed to the reform. These findings also shed light on one of the multiple channels through which dictatorships can alter individuals' preferences and values (Brum, 2018).

Focusing on subjective well-being is another major departure from the previous literature on school segregation, which has looked at outcomes such as educational attainment (Guryan, 2004; Johnston et al., 2007; Reber, 2010; Billings et al., 2014) and crime (Kling et al., 2005: Weiner et al., 2009; Billings et al., 2014). Echenique et al. (2006) find that school racial segregation is negatively associated with the likelihood of inter-ethnic couples but uncorrelated with subjective well-being. The choice of this measure of well-being is especially useful as it can reflect subtle changes in living conditions of individuals that would otherwise pass unnoticed.

Recent research has assessed the long-term impact of moving to richer neighborhoods on aspects such as health, subjective well-being and earnings, taking advantage of a major social experiment in the US, the Moving to Opportunity Program (Sanbonmatsu et al., 2012; Ludwig et al. 2012;

Chetty et al., 2015). Despite relying on a randomized intervention, these studies have reached mixed conclusions and lack general equilibrium considerations. My results complement this work by looking at nation-wide reform that changed the pool of peers to which children were exposed to by fostering the entry of thousands of private schools across the country.

While there has been extensive research analyzing the effects of the 1981 voucher reform in Chile, its focus has been on its short-run impact on educational quality and segregation.[2] These studies have reached mixed conclusions regarding the impact on students' performance (Hiseh and Urquiola, 2006, Contreras et al., 2012) but there is a consensus that the increase in subsidized private enrolment induced by the the reform led to a significant rise in school segregation (Hiseh and Urquiola, 2006; McEwan, Urquiola and Vegas, 2008; Elacqua, 2012; Valenzuela, Bellei and De los Rios, 2014; Santos and Elacqua, 2016). I contribute to this literature by examining for the first time the long-term consequences of the 1981 voucher reform on a variety of adult outcomes.

The remainder of this paper is structured as follows. Section 2 describes the historical background, focusing on the main changes introduced by the reform and why it led to a rise in segregation. Section 3 and 4 detail the data sources and the identification strategy respectively. Section 5 presents the results, discusses potential mechanisms and provides some robustness checks. Finally, Section 6 summarizes the findings and draws some policy implications.

2   Historical Background

In this section I discuss the main changes introduced by the 1981 educational reform and its consequences. In a nutshell, a new school choice and financing system was implemented under which children could attend any public or subsidized private school without having to pay any fees. Schools were funded with a flat-per-student subsidy (voucher) and also granted greater administrative flexibility. As a result, in just five years the number of subsidized private schools doubled and the share of students they enrolled increased from 15 to 30 percent. Since the exodus from public schools was characterized by middle-class students, the privatization process induced by the vouchers led to a sharp increase in school segregation.

---

[2]There is some recent work examining the impact of Pinochet's privatization policy on the development of business corporations (González, Prem and Urzúa, 2020).

2.1 Pre-Reform Years: 1950-1981

In the period 1950-1973, the main objective of educational policy in Chile was the expansion of access. There were large investments in new establishments, resources, and a double-journey system was introduced.[3] Enrollment in basic education increased from 70.7 percent in 1947 to more than 97 percent in 1970 (Cruz, 1999; Cox, 2003). Immediately after the 1973 Military coup, a cleansing of teachers and school heads took place, expelling those who differed ideologically from the regime (Echeverria and Hevia, 1981). Teachers' unions were dismantled and the curriculum was also slightly modified to incorporate national and traditional values (Cox, 2003). Enrollment in secondary education suffered a sudden stop, growing at only 3.4 percent per year in 1973-1979 compared to 15.4 percent in 1967-1973. Besides these aspects, there were no significant changes in the educational system before 1981.

At this time, there were three types of schools:

- **Public schools.** They accounted for 70 to 80 percent of total enrollment in the 1960s and 1970s. They were financed through annual transfers from the Ministry of Education and did not charge fees. The amount of these transfers was determined from previous costs, the number of students and the characteristics of the school (Hsieh and Urquiola, 2006). Education was of similar quality across public schools since they all followed the same curriculum.

- **Subsidized private schools.** They accounted for around 15 to 20 percent of enrollment in this period. More than half were run by the Catholic Church and the rest by Protestant churches or non-religious foundations. They did not charge fees and were not for-profit (Espinola, 1993). Although they received transfers from the Ministry of Education they were severely underfunded and had to largely rely on their own resources.[4] Most followed the same curriculum as public schools, often including additional religious content.

- **Fee-charging private schools.** These were elite institutions offering specialized and high-quality education. They were run for-profit, without any control from the Ministry of Education, and did not receive any public funding. Fee-charging private schools had admissions requirements, such as parental interviews or entry examinations.

---

[3]Besides standard morning classes, some groups started attending school in the afternoon, thus doubling the capacity of public schools (Cruz, 1999).

[4]Not only were the level of the subsidies insufficient to cover all costs, but the delay in monthly payments further eroded their value due to inflation (Hiseh and Urquiola, 2006).

Prior to the reform, both public and subsidized private schools operated similarly and offered education of comparable quality. They did not charge fees nor did they select, thus generating a mixed pool of children from all social classes. Besides the wealthiest families who sent their children to fee-charging private schools, most children attended the closest school to their homes regardless of whether it was public or subsidized private (Hsieh and Urquiola, 2006).

2.2 The 1981 educational reform

The Military Government aimed to liberalize the educational system with the objective of making it more efficient. It believed that higher flexibility and competition for students among schools would improve educational outcomes while reducing the costs. The 1981 reform targeted these aspects by decentralizing the administration of public schools and introducing a system of school vouchers.

1. Decentralization and flexibility. The decentralization process implied the transfer of all public schools from the Ministry of Education to municipalities in the first months of 1981. This meant that local governments had to administer all financial, material and human resources of schools (Espinola, 1993). Decentralization did not imply a loss of control, as municipality mayors were appointed by the regime, and these appointed school heads (Cox, 2002). The Ministry of Education kept a subsidiary role of control and supervision and suffered a large reduction in its personnel, falling from 20 thousand employees in 1982 to around 3 thousand in 1989 (Prawda, 1992). Public teachers stopped being civil servants to become private employees under standard contracts with municipalities and although they could be fired freely this rarely happened (Carnoy, 1998).[5] Schools were allowed to choose between a 30- or 25-hour week and reallocate hours between subjects. In the last years of secondary education students were also allowed to choose between subjects (Espinola, 1990). Since there were no longer content guides or methodological plans, teachers were allowed to implement the practices they considered appropriate (Espinola, 1993). Most public schools however continued to follow national guidelines and barely changed their curriculum (Cox, 2002). The greater flexibility often meant fewer hours and less content in the poorest context (Espínola, 1993).

2. Vouchers and school financing. The crucial change the 1981 reform brought about was the in-

---

[5]Recall that Teachers' Unions had already been dismantled in 1973.

troduction of a flat-per-student subsidy (i.e., voucher) for public and subsidized private schools, entirely replacing the previous financing system. Under the new voucher scheme, children could attend any public or subsidized private school regardless of its location and without having to pay any fees. Schools would receive a monthly payment proportional to the number of students enrolled, with an adjustment for average attendance (Carnoy and McEwan, 2002; Espinola, 1993). Importantly, the value of the per-student subsidy was the same for both public and subsidized private schools, and did not take into account the socio-economic status of students (Elacqua, 2012). All previously-subsidized private schools (e.g., religious schools) were automatically enrolled in the voucher system, while many new ones entered the market and joined the system.[6] While fee-charging private schools could join the voucher scheme they rarely did so, as that would have implied losing their ability to levy fees. The conditions to receive the subsidies were: (i) not to charge fees, (ii) have classes of at most 45 students, (iii) follow general guidelines concerning the curriculum and (iv) abide by some infrastructure regulations (Espinola, 1993; Mizala and Romaguera, 2000). To incentivize the creation of new subsidized private schools, the value of the subsidy was set at 30 percent above the average spending per student in 1980 and 61 percent higher than the average per student transfer that subsidized schools use to receive before the reform (Espinola, 1993).

2.3   Privatization of education and rising segregation due to the 1981 reform

The extremely favorable conditions of the voucher scheme in terms of financing encouraged new private schools to enter the market and join the system. As shown in Figure 1, close to 15 hundred new subsidized private schools opened between 1980 and 1985, doubling the existing stock of these type of schools. According to Hsieh and Urquiola (2006), 84 percent of these new private schools were for-profit institutions. Likewise, there was a large increase in subsidized private enrollment following the reform. The share of students attending subsidized private schools rocketed from 15 percent in 1981 to 33 percent in 1988, as illustrated in Figure 2. Regarding fee-charging private schools, both the total number of establishments and their enrollment remained fairly constant.

The exodus out of public schools was dominated by middle-class students, thus concentrating children from low socio-economic status in these schools (Hsieh and Urquiola, 2006). According

---

[6]I will refer to all private schools receiving the subsidy as subsidized or subsidized private schools irrespective of whether they were already present before the reform.

to a number of studies, the 1981 voucher reform seem to have had a negligible effect on educational quality measured but raised significantly school segregation and inequalities between schools (Auguste and Valenzuela, 2004; Hsieh and Urquiola, 2006; McEwan, Uquiola and Vegas, 2008; Valenzuela, Bellei and De los Rios, 2013).

The strong positive relationship between family income and children's enrollment in private schools is depicted in Figure 3. In the bottom income decile more than 80 percent of children attended public schools. By contrast, in the top decile 80 percent of children were enrolled in private schools. Subsidized private enrollment increases linearly with income between the first and the seventh decile, going from 15 to 35 percent. As middle-class children were much more likely to attend subsidized private schools than poorer ones, the long-term impact of the reform may differ substantially depending on the family background of the children exposed.

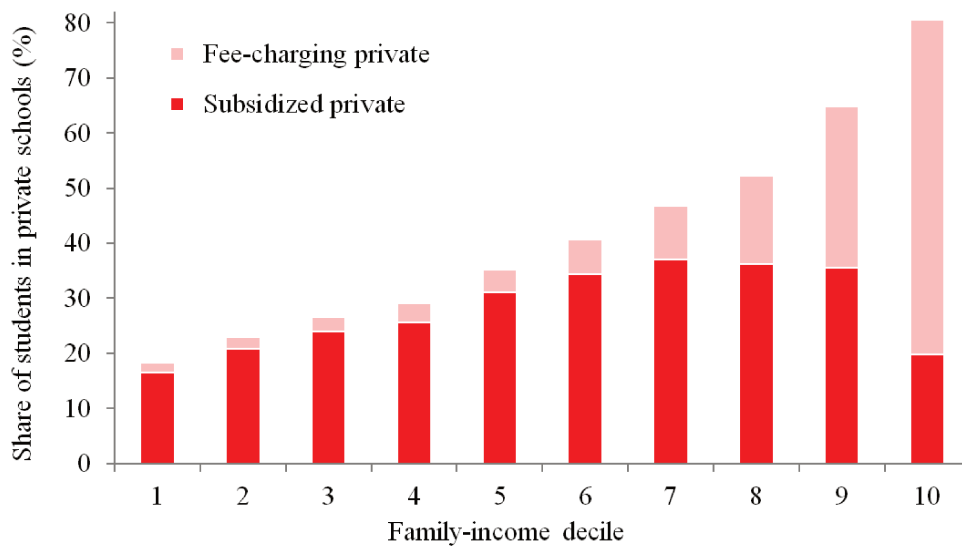Figure 1: Total number of establishments by school type, 1980-1995
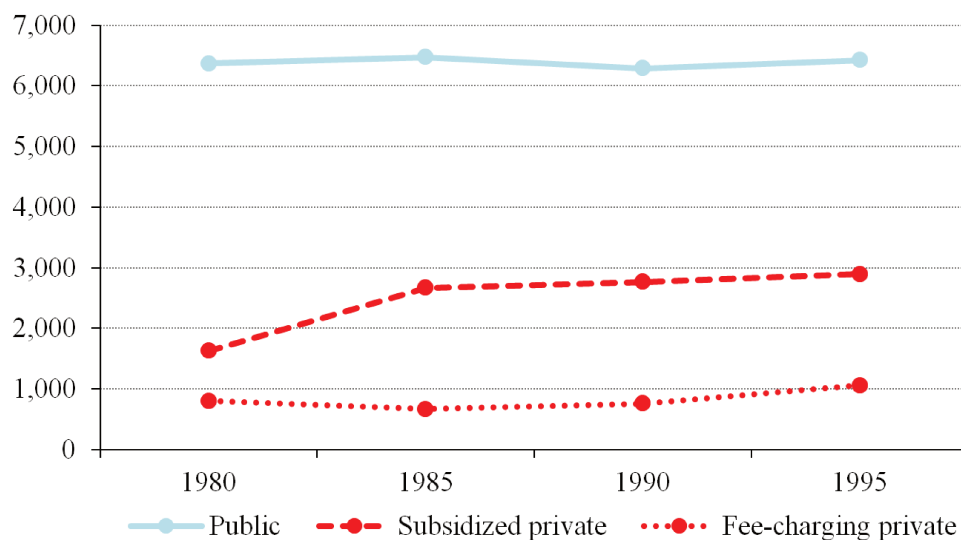
Note: Data from the Ministry of Education. All primary and secondary schools included.

Figure 2: Total enrollment by school type, 1970-1995
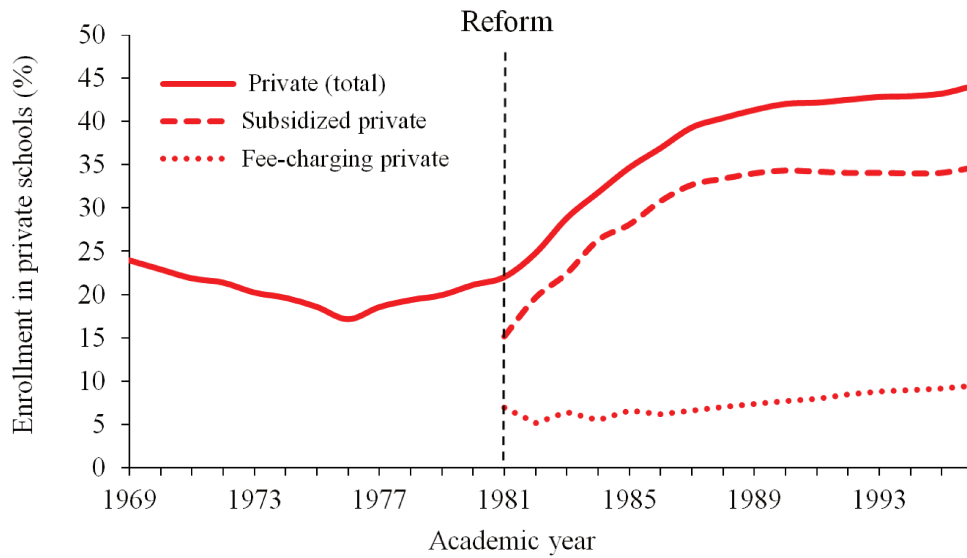
Note: Data from the Ministry of Education. All primary and secondary schools included.

Figure 3: Enrollment in private schools by family income, 1990-1992

Note: Own elaboration using household survey CASEN 1990 and 1992.

As mentioned, the rise in private schools came hand in hand with increasing segregation. Figure 4 shows the relationship between the share of students attending subsidized private schools in a province and the relative income of families with their children attending public rather than subsidized private schools. There is a strong negative relationship, implying that the higher is subsidized private enrollment in a province, the greater is stratification across schools. On average, a 10 percentage-point rise in the share of students attending subsidized private schools in a province is associated with a widening of gap in family income between public and subsidized private schools by 4 percentage points. Data at the more disagregated level points in the same direction. McEwan, Urquiola and Vegas (2008) use district level data and show that when the extent of privatization was low, families tended to have similar incomes regardless of the type of school of their kids. However, the greater the enrollment in subsidized private schools, the lower the relative income of public-school families relative to subsidez ones in those districts.

The rise in private enrollment led to greater stratification not only because relatively better students left public schools but also because within-sector stratification is higher among private schools. Of all schools, public schools have the most diverse student body in terms of socio-economic status and indigenous origins (Elacqua, 2012; Valenzuela, Bellei and De los Rios, 2014). Similarly, Bravo, Contreras and Sanhueza (1999) find that public schools are much more homogenous than subsidized private schools in terms of performance. For instance, the share of the variation in test scores that is explained by differences between schools within a sector is 37 percent smaller for public schools than subsidized private schools.

2.4   Why did so few low-income children attend subsidized private schools?

The 1981 reform led to the creation of more than a thousand subsidized private schools in five years, which did not charge fees. Yet, very few low-income children attended these schools. There are least three reasons behind this striking fact: i) the strategic location of schools, ii) parents' beliefs and choice behavior, and iii) school selection procedures.

There is evidence that the new subsidized private schools were more likely to locate in richer municipalities and within municipalities, in relatively better-off neighborhoods with a larger presence of medium and upper-class families (Mizala and Romaguera, 2000; Bellei, 2009). In 1988, 28 percent of municipalities did not have a single subsidized private school; most of them rural. Even

Figure 4: Subsidized private enrollment and stratification across schools, 1992
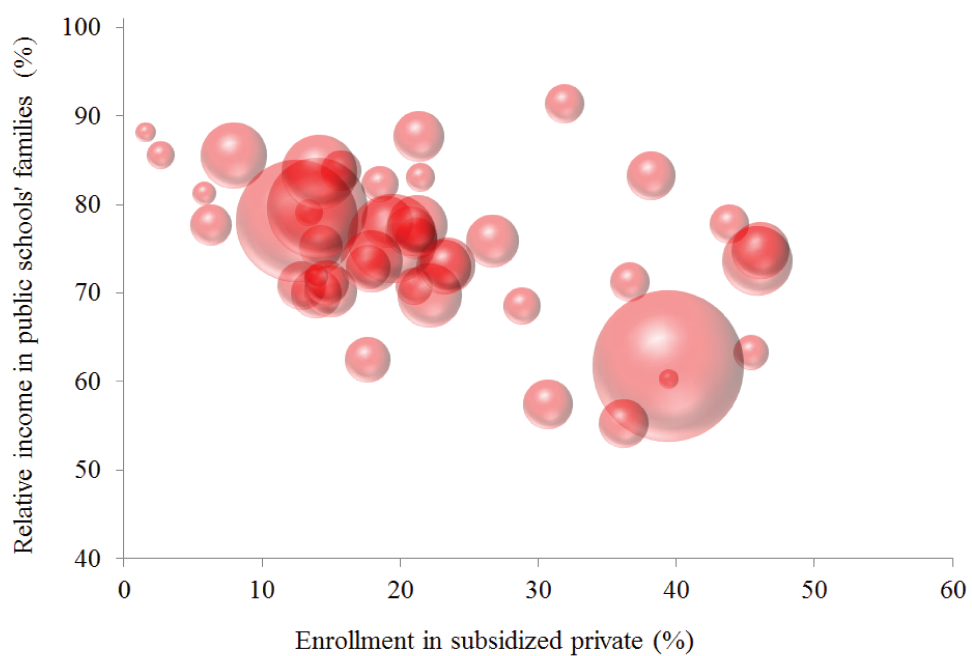
Note: The figure shows the share of students attending subsidised private schools in a province in 1991 and the ratio of household income between families with kids attending public schools rather than subsidized private ones. The size of thee dots indicates the population of the province. Data from household survey CASEN 1992 and the Ministry of Education.

when there was a subsidized private school in town, their strategic location in wealthier areas may have hindered the enrollment of poorer children due to transportation costs, thus exacerbating stratification. Survey data confirms that even today, the main reason why parents do not consider sending their children to subsidized private schools is because there are none nearby (Mediavilla and Zancajo, 2015).

Lower-educated parents could be less likely to send their children to higher-quality schools because they do not have information about how good the schools are or because they put a lower value on education quality. Although in the early 1980s the Military Government introduced a standardized examination for all schools, the results were never made public and therefore parents could not judge schools on the basis of their quality.[7]. Espinola (1993) documents that new subsidized private schools tried to attract higher socio-economic status families by endowing themselves with symbols that were previously associated with elite private schools such as uniforms or English names. Other research in the context of Chile shows that the lower parental education and income are, the greater the probability that children attend the closest school irrespective of local availability (Chumacero, Gomez and Paredes, 2011). According to Mediavilla and Zancajo (2015), parents with children in public schools are 63 percent more likely to cite proximity as a relevant choice factor compared to those in subsidized private schools. There is additional evidence showing that most parents, irrespective of their educational level, do put considerable value on academic quality when exercising choice (Schneider, Elacqua and Buckley, 2006). However, they include in their choice set schools with very different performances in test scores, which contradicts their stated preferences. Interestingly, schools in the choice set barely differ in terms of the socio-economic composition of the students. These findings lead Schneider, Elacqua and Burckley (2006) to conclude that the most important choice factor underlying parents' actual decisions is the socio-economic status of peers.

Although subsidized private schools were not allowed to charge fees under the voucher system, they skimmed-off students via various selection procedures. Some of them required income or

---

[7]In 1982 the Ministry of Education created a national examination called PER (Programa de Evaluacion de Rendimiento Escolar) with the objective of evaluating the quality of education. The PER was implemented in all public and most subsidized private schools. The fierce opposition of teachers together with the high costs involved led to its disappearance after 1984. Nevertheless, a similar test was launched in 1988 under the name of SIMCE (Sistema de Medición de la Calidad de la Educación) and has been running to date with subsequent improvements and expansions. Today parents have access to detailed information about the performance of their children's school and others in their municipality.

religious certificates, held interviews with parents or even had entry examinations. For instance, Contreras, Sepulveda and Bustos (2010) find that 56 percent of children attending subsidized private schools were subject to some form of selection compared to only 5 percent of students in public schools. Moreover, the higher the socio-economic status of children, the greater the probability of having been through a selection process. For example, students from the top income decile attending subsidized schools were subject to some form of selection in 74 percent of cases (compared to 25 percent for top income students in public schools). In recent years, public schools increasingly apply selection criteria and mimic the practices of private schools (Contreras, Sepulveda and Bustos, 2010; Portales and Heigil, 2015). An additional factor of stratification is that subsidized private schools had the right to expel students for behavioral and academic reasons, while public schools were not allowed to do so (Bellei, 2009; Elacqua, 2012). Schneider, Elacqua and Buckley (2006) find that only around half of the parents consider more than one school in their choice set, which may reflect these barriers.

In sum, low-income children were much less likely to attend subsidized private schools because they were often far from them, and even if not, schools could eventually refused the admission to these children or expelled then even after having been admitted.

3 Data

This section describes the main data sources used in the analysis. The main data for the outcomes of interest comes from the National Household Survey of Chile (CASEN), which includes information on subjective well-being and a number of socio-demographic factors. The historical enrollment data used to construct a measure of exposure to the reform comes from the archives of the Ministry of Education. These data is linked to each individual in the CASEN based on their municipality of birth. Additional provincial and municipality level data are obtained from different sources and summarized in Table A.1.

3.1 Household survey data

The main data comes from the National Household Survey of Chile (hereafter CASEN). This survey is conducted by the Ministry of Social Development and Planning every couple of years to assess the living conditions of the population, evaluate public policies and compute official measures of inequality and poverty. The first waves in the 1990s did not include all municipalities.

Today, the survey is designed to provide accurate estimates at the local level. CASEN contains information on demographics, education, labor-market outcomes, income and health, among others.[8] Retrospective information about parents' education is available since 2009. In addition, the 2011 and 2013 waves, which are the two years used in the analysis, include a standard life-satisfaction question on a 1 to 10 scale. The precise wording is: "Taking into consideration all aspects of your life, how satisfied are you with your life these days? From 1 to 10 where 1 means completely unsatisfied and 10 completely satisfied´´.

All individuals with no missing information on life satisfaction and parents' education are included in the analysis. This leads to a nationally-representative sample of 58,825 individuals. Table A.2 in the Appendix shows the descriptive statistics of the main variables in the full sample as well for different categories of family background based on parents' education. The over-representation of women can be explained by their remarkably lower participation in the labor force and therefore greater likelihood of being in the household at the moment of the interview.

3.2 Enrollment records

Historical enrollment data was obtained and digitized from archives of the Ministry of Education of Chile. For every of the 346 municipalities there is information on the total number of students enrolled in public, subsidized private and fee-charging private schools for each grade and year since 1980. Given the poor state of the physical documents the data had to be digitized manually. Three years were selected; one prior to the reform (1981),[9] one when the rise in private enrollment stopped (1988) and an additional third year for additional analysis (1991).[10] Using this data, I construct a measure of 1980s privatization as the rise in the share of students enrolled in subsidized private schools between 1981 and 1988 in each municipality.[11] Table A.3 in the Appendix presents some descriptive statistics of the enrollment data from the original sources (*municipality level*) and once it has been linked to the individuals from the household survey (*individual level*).

---

[8]Certain waves incorporate supplementary questions about concrete themes, such as the quality of neighborhoods, networks or gender issues.

[9]The enrolment data is from the academic year 1980/1981, while the voucher system was in place from the academic year 1981/1982 on wards.

[10]Similar data was used by Hsieh and Urquiola (2006) for the years 1981 an 1988 although it was lost afterwards.

[11]This comprises all students in primary and secondary education. Since 1965, primary education ("básica´´) lasts for eight years and secondary ("media´´) only four.

3.3 Supplementary data

I rely on supplementary data coming from multiple sources. These include measures of human capital, inequality, employment, indigenous populations, urbanization, population density, religiosity, mortality, political orientation, political repression and number of schools, among others. The information is either at the provincial or local level. For instance, I use historical data from the 1970 and 1980 censuses to control for underlying trends in demographic and economic factors that could be correlated with both the degree of privatization in the 1980s and subjective well-being today. I provide further details in the *Empirical Strategy* and *Robustness Checks* sections. All data sources are described in Table A.1 in the Appendix.

4   Empirical strategy

The goal of this paper is to assess the long-term impact of the 1981 voucher reform on subjective well-being. To do so I exploit the sudden increase in subsidized private enrolment following the reform and its considerable heterogeneity across municipalities. Since higher enrollment in subsidized private schools translated into greater segregation, this measure serves as good proxy in the absence of more detailed data.[12]

A straight-forward way of examining the long-term impact of the reform would be to look at the relationship between the increase in private enrolment in the 1980s and the current life satisfaction of the individuals who attended school those years. Note however that the expansion of subsidized private schools during the 1980s is potentially endogenous, that is to say, it could be correlated with other factors that also affect life satisfaction. For instance, we can suppose that enrollment in subsidized private schools grew faster in richer towns. To the extent that these towns are still relatively rich today, we could erroneously conclude that the reform had a positive impact on 14 well-being. Since we cannot control for all local factors that may confound the effect of the reform, the naive cross-sectional approach is problematic. To tackle this issue the identification strategy compares individuals exposed to the reform to others not exposed in the same area (the exposure dimension –E-). By doing so any time-invariant local effect is difference out. The identification considers individuals born in the early 1970s against others born in the early 1960s. The younger (treated) group was in school after the reform took place and hence spent some of

---

[12]It is not possible to construct indices of segregation for this period as there is no information about students' characteristics at the school or local level.

their education years in a more segregated environment. On the contrary, individuals in the older (. control) group did not experience the changes induced by the reform as they had already left school by 1981.[13] Given that the identification relies on the comparison of individuals born in different years, I introduce cohort fixed effects to control for any year-specific effect on well-being. The key explanatory variable is the increase in subsidized private enrollment between 1981 and 1988 in each municipality (the intensity dimension -T-). As shown in Figure A.4 in the Appendix, subsidized private enrolment grew sharply in the first years after the reform but remained flat after 1988. Each individual is linked to the rise in subsidized private enrolment in their birth municipality.[14] By introducing an interaction between the rise in subsidized private enrolment and the exposure measure (instead of the rise in private enrollment alone) we can introduce municipality fixed effects to capture all of the time-invariant local factors correlated with well-being.

There is a possibility that the increase in subsidized private enrolment was correlated with other unobserved trends that have long-lasting impacts on well-being. For instance, it may be that subsidized private schools opened more frequently in places where schooling was growing faster, so that comparing younger to older cohorts may mistakenly suggest a positive impact of the reform on well-being. To limit the concern of confounding trends I use census data to control for the evolution of population density, urbanization, schooling, unemployment and the share of indigenous people over time. These measures vary thus across municipalities and cohorts,[15] and although not exhaustive, provide further strength to the identification. It is hard to believe that unobserved trends are behind the results as they have to be strongly correlated with the rise in subsidized enrolment and to have a persistent impact. I show later that controlling for the trends that I do observe makes no difference to the results.

Figure A.5 in the Appendix illustrates the approach. Children who left school before the academic year 1981-1982 were not exposed to the reform. Between 1981 and 1988 new private schools entered the market, enrollment in subsidized private schools increased substantially and at the same

---

[13]Individuals in the treatment group are on average 8 years old in 1982 while those in the control group are on average 18. In the next section I modify the cohort windows considered in each group and carry out the analysis as if the reform had occurred in a different year

[14]It is possible that some attended schools in other municipalities but, due to the low levels of internal migration at young ages, these are probably only negligible. According to the CASEN, the share of children (< 18 years old) living in the municipality in which they were born is very high. It fell from 83 percent in 2009 to 79 percent in 2015, which may suggest a higher figure in the 1980s.

[15]For individuals born in the 1960s I use data from the 1982 census and for those born in the 1970s I use the 1992 census.

time so did segregation. Therefore, children attending school during these years were partially exposed as the process of privatization was ongoing. In addition to rising segregation, in this period there could have been some disruptive effects because some students and teachers switched schools. All children in education after 1987 were exposed to the highest degree of segregation as after 1988 subsidized private enrollment stopped rising.

Formally, the identification strategy consists in estimating the following equation by OLS:

$$y_{icj} = \lambda_j + \lambda_c + \delta E_{ic} + \rho T_j \times E_{ic} + X_{icj}\Theta + Z_j\Phi + \epsilon_{icj} \tag{1}$$

where y_{icj} standardized measure of adult life satisfaction for individual i in municipality j and cohort c, λ_j and λ_c are municipality and cohort fixed effects respectively , E_{ic} is a dummy for the individual being born between 1972 and 1976 and zero if she was born between 1962 and 1966, T_j is the increase in subsidized private enrollment between 1981 and 1988 in municipality j, X_{icj} is a vector of individual characteristics, Z_j are municipality-cohort trends, an ϵ_{icj} are individual-specific error term clustered at the municipality-cohort level. The interaction coefficient $T_j \times E_{ic}$ should be interpreted as the impact of one percentage point increase in subsidized private enrollment in the 1980s on adult life satisfaction in 2011-2013.

Note that there is no information about the type of school individuals attended in their childhood. This is problematic as we may expect that the reform affected students differently precisely according to whether they stayed in a public school with worse peers or went to a subsidized private school with better peers. I circumvent this issue by distinguishing individuals regarding their family background. As shown in Figure 3, enrolment in subsidized private schools is considerably lower in the bottom income deciles than in the middle of the income distribution. Although I do not have information on parents' income, I use parents' years of schooling to construct a three-level measure of family background. Given that parental education varies significantly across municipalities and over time, I construct a measure that takes these two aspects into account. Concretely, for a given cohort t, I consider the parental schooling distribution of all individuals born between t and $t + 4$ in their respective province. I then categorized individuals according to their family background:

1. Low family background if parental schooling is in the bottom 4 deciles of the distribution.

2. Medium family background if parental schooling is in the 5th to 8th deciles.

3. High family background if parental schooling is in the top 2 deciles of the distribution.

I carry out all of the analyses distinguishing by family background. Most low family-background children continued attending public schools after the reform, with subsidized private schools attracting middle-class students they did so surrounded by peers of lower socio-economic status on average. Among medium family background students, some moved to subsidized private schools after the reform and hence had better peers on average while others remained in public schools therefore in a worse environment than if the reform had not occurred. Similarly, among high family background students many kept attending fee-charging private schools while others moved from public to subsidized private schools and thus were surrounded by better peers. As a result, the increase in subsidized private enrolment measures the rise in segregation and at the same time the extent to which the socio-economic status of school peers changed.[16]

The identification strategy relies on a number of assumptions. First, the exposure measure (E_c) assumes that children in the control group (born between 1962 and 1966) were not affected by the reform, as they had left school by 1981/1982 and because segregation was not already rising beforehand. However, it is unfortunately not possible to test whether segregation had increased before or evaluate its level as there is no information on students' characteristics within municipalities. Likewise, the hypothesis of spillovers it is difficult to test: if there was some "contamination" of the control group then the estimates will represent a lower-bound of the true effect. A second assumption is that segregation did not continue increasing immediately after 1988, otherwise the treatment variable will be measured with error. This assumption reflects data limitations as I only have enrollment data across municipalities for 1981, 1988 and 1991. However, at least at the national level, subsidized private enrollment barely increased after 1988 as shown in Figure A.4 in the Appendix. In the Robustness Checks section I also use the interval 1981-1991 instead of 1981-1988. There are two subtle assumptions regarding homogeneous effects. First, it is assumed that a given increase in private enrollment has the same impact irrespective of the year it occurs,

---

[16]As a robustness, I change the definition of family background in different ways. For instance, I alter the thresholds which define each group and also consider national-level cutoffs (i.e., independent of the level of schooling in each region).

in other words, independently of the age of individuals. Second, it is assumed that the exposure is the same for individuals born in the same cohort irrespective of much time they spent in school. All in all, these assumptions imply that the estimates I provide represent an average treatment effect for the whole population, and not for those actually affected in a single direction or a subpopulation of them. For example, some individuals of medium family background attended public schools and others subsidized private schools. Within this group there are individuals of different ages affected differently and also staying in education for different durations, all of which implies measurement error. The identification strategy has the advantage of being conservative and conclusive about the overall impact of the reform but the disadvantage that the estimates will be biased toward zero and hence more difficult to identity and attribute to a concrete population. If we find any significant impact, we can presume that the true impact for the population actually affected is even larger.

5 Results

I first discuss some graphical evidence that supports the empirical strategy of the paper. I then estimate the impact of 1981 school privatization on current life satisfaction using equation (1) and distinguishing by family background. Next, I modify the cohort windows considered in the treatment and control groups in multiple ways to test the robustness of the results. I also analyze heterogeneous effects.

5.1 Graphical evidence of common trends

The empirical approach adopted relies on the crucial assumption that the increase in subsidized private enrollment during the 1980s is not correlated with any other omitted trend making life satisfaction differ across cohorts within the same municipality once those individuals are adults. In the spirit of a difference-in-difference framework, treatment and control areas should have similar common trends in other dimensions prior to the policy change. Similarly, the evolution of life satisfaction across individuals not affected by the policy should be similar regardless of the location. I will provide evidence in fact showing that the relationship between year of birth and life satisfaction is the same in areas of high and low privatization for individuals who left school before the reform in 1981.

Panel A focuses on average life satisfaction while Panel B looks at happiness inequality measured

by the Gini coefficient. The two graphs on the left separate municipalities below and above the median increase in subsidized private enrollment in the 1980s while the two on the right compare municipalities where there was zero increase and others in which the rise was above 20 percentage points.Looking first at Panel A, both figures confirm that for all cohorts average life satisfaction levels are not statistically different in municipalities with low and high privatization. There is a general positive trend indicating that younger individuals tend to be happier than the older, peaking in the late 30s, in this case individuals born in the late 1970s. Although the 95 percent confidence intervals for the two groups overlap, municipalities with more privatization have a slightly higher level of life satisfaction on average. This gap, however, vanishes by 1976 which it is a first indication that something may have happened in that period. One can formally test whether the pre-trends are equal by applying the identification strategy set out in equation (1) to older cohorts. I provide further evidence in this respect in a following section.

Panel B shows that inequality of life satisfaction falls for the younger cohorts. While the U-shaped relationship between age and life satisfaction has been well documented, this positive relationship between happiness inequality and age has not received much attention. Note that for those born in the 1960s the *gini* coefficient is persistently larger in municipalities that experienced greater privatization. However, the correlation coefficient between current happiness inequality for cohorts born in the 1960s and privatization during the 1980s is only 0.035. For individuals born in the 1970s and 1980s there are no significant differences in happiness inequality according to the degree of privatization.

5.2   The long-term effects of the 1981 education reform

Table 1 shows the main results from estimating equation (1) in columns 1 to 4. Remember that the coefficient of interest is the interaction between the change in subsidized private enrollment in 1981-1988 and the exposure variable. All regressions include a set of individual controls, cohort fixed effects, municipality fixed effects and municipality trends.

Overall, the analysis suggests little effect of the privatization reform on average life satisfaction. In column 1, the interaction coefficient is close to zero and far from significant. However, this aggregate result masks substantial differences. In columns 2 and 4, individuals from low family backgrounds seem to have been harmed by privatization, while the opposite holds for those

from high family backgrounds. Individuals from medium family background do not seem to have been affected on average. This may however be the net result from both positively- and negatively-affected individuals.

Table 1: Effect of 1981 voucher reform on adult life satisfaction (2011-13)

	Life satisfaction (standardized)			
	All	Low FB	Medium FB	High FB
	(1)	(2)	(3)	(4)
Exposed to the reform	-0.0578	-0.4819	0.3300	0.1990
(born in 1972-1976 rather than 1962-1966)	(0.4101)	(0.6708)	(0.7495)	(0.7657)
Δ Private enrollment (%) * Esposed to the reform	-0.0008	-0.0059***	-0.0013	0.0038*
	(0.0010)	(0.0018)	(0.0019)	(0.0021)
Individual controls	Yes	Yes	Yes	Yes
Cohort Fixed Effects	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes
Municipality Trends	Yes	Yes	Yes	Yes

Note: Columns 1 to 4 estimate the model in equation (1). The dependent variable is a standardized measure of life satisfaction. FB stands for family background based on parents' education. Standard errors clustered at the municipality-decade level in parenthesis. Statistical significance: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Individual controls include a dummy for being women, from aboriginal origin, living with both parents during childhood, parental schooling, and age and age-squared. The municipality trends include the evolution of population density, share of children, urbanization, employment, average schooling and schooling inequality from 1970 to 1982.

The estimate of -0.0059 in column 2 implies that among those from a low family background, being exposed to a rise in subsidized private enrollment of 10 points (approximate one sd.) during childhood is associated with a fall in long-term life satisfaction of 5.9 percent of a standard deviation. This impact is equivalent to the effect of lower household income of around 30 percent or reducing education by 3 years. The analogous figure for those is a rise in long-term life satisfaction of 3.8 percentage points of a standard deviation. The coefficient in this case is much less precisely estimated. Given this differential effect of the reform depending on family background, it is unsurprising that the overall net impact is not significantly different from zero. We should also bear in mind that this is a point-year effect, so that low family-background individuals exposed to higher segregation as children may have been systematically worse off year after year during their adulthood. It is not possible to determine when this negative effect starts to be seen, however, if present even for only a few years it can represent a substantial welfare loss. Note that coefficient on the dummy "Exposed to the reform´´ is far from significant in all the regressions, so

that there are no systematic differences in life satisfaction between cohorts born in the early 1960s and in the early 1970s. This finding reduces any concerns that the year of birth may play a role in the results.

Alternative measures of exposure. Rather than comparing exposed to not-exposed cohorts, I perform an additional exercise using alternative measures of exposure. Given that the reform took place in 1981 but private enrollment rose until 1988, I consider the length of exposure under three different scenarios: (i) years of compulsory education after 1981, (ii) years of compulsory education after 1988, (iii) years of compulsory education after 1988 count as full exposure and the years during 1981-1988 count partially (e.g., as if exposure grew linearly between 0 and 1 in 1981-1988).

The results are summarized in Table 4 in the Appendix. As in the main analysis, only individuals from lower family background seem to have been affected by the reform, I would hence restrict the discussion to this group. Column (1) shows that considering full exposure as attending school after 1981, a 10 percentage point increase in private enrollment would lower adult life satisfaction by 0.7% for each year of exposure. In other words, if an individual spent all her basic education (8 years) under the more privatize system, her life satisfaction as an adult would be expected to be 5.6% of a standard deviation lower. Column (2) shows that considering full exposure as attending school after 1988, the impact is much higher. This result makes sense as previously mentioned, the levels of school privatization and segregation were much higher in 1988 than in 1981. In this case, a 10 percentage point increase in private enrollment would lower adult life satisfaction by 4.2% for each year of exposure. This implies that a attending 8 years of schooling after the peak in privatization (1988) rather than no exposure at all could reduce adult life satisfaction by as much as 33.6% of a standard deviation. Finally, column (3) presents a middle-point between both scenarios, in which the schooling years between 1981 and 1988 due to contribute to lower subjective well-being.

5.3   Testing for unobserved trends

I carry out three analyses to test for unobserved municipality trends that are correlated with the rise in private enrollment in the 1980s and life satisfaction across cohorts. I first perform a falsification test, re-estimating equation (1) for different cohort windows as if the reform had occurred in another year. We would here expect to find significant effects only in the vicinity of the true

year. Second, I set the control group to older individuals and estimate equation (1) with treatment groups of different ages. In this case, only people born in the 1970s or later should be affected, as they were the only ones to be exposed. Third, I estimate equation (1) including in the control and treatment group individuals of very similar ages and compare all the possible cohort windows. In this test I expect no significant differences between the treatment and control group, with the exception being some cohorts were actually exposed while others were not.

Panel A in Figure 5 plots the interaction coefficient in equation (1) for each potential year that the reform could have occurred. The estimate of year t is obtained by setting as the control group individuals born between $t - 20$ and $t - 16$ and as the treated group individuals born between $t - 10$ and $t - 6$ For instance, the estimate of 1982 comes from regressing equation (1) with individuals born in the period 1972-1976 (i.e., treated, $E_{ic} = 1$) and born in the period 1962-1966 (i.e., control, $E_{ic} = 0$), while the estimate of 1981 comes from decreasing in the two windows by one year. In the case of low family background, the results show that the interaction coefficient is negative, large and significant only in the early 1980s. The coefficient moves between -0.005 and -0.006, implying that a 10 percentage point increase in subsidized private enrollment in the 1980s is associated with a fall in adult life satisfaction of around 5 percentage points of a standard deviation. For medium family background the estimates are not significant but do display a negative trend starting in the early 1980s. The estimates for high family background are negative and slightly significant for the mid-1970s and positive and slightly significant for the late 1980s. Overall, the figures suggest that unobservable trends are not driving the results, as the estimate are only significant when setting the reform at the beginning of the 1980s, the date it actually occurred.

The second analysis sets as the control group ($E_{ic} = 0$) individuals born in 1950-1957, who should be completely unaffected by the reform. I estimate equation (1) for every year t, defining as the treatment group ($E_{ic} = 1$) individuals born between t and $t + 4$. Panel B shows the interaction coefficient for each of these estimations. The estimates are insignificant for cohorts born in the late 1950s and 1960s, regardless of family background. This suggests that the findings are not driven by a spurious correlation between an unobserved trend correlated both with long-run happiness and the penetration of private schools in the 1980s. Focusing on low family-background individuals, those born in the 1970s in municipalities that would experience a larger privatization are considerably less satisfied today than their older counterparts in that same municipality. The esti-

mated coefficients again fluctuate around -0.006, meaning that being exposure to a 10 percentage point rise in subsidized private schooling is associated with a reduction in adult life satisfaction of 6 percent of standard deviation. For medium family-background individuals the impact of the reform seems to be delayed: only those born in the late 1970s exhibit lower levels of life satisfaction as private enrollment rose in their municipality. The estimates for high family background individuals are always insignificant, although for the youngest cohorts the coefficients are positive, of moderate size and close to significant.

The third approach comparing adjacent cohort windows has the disadvantage that individuals in the treatment and control groups were similarly exposed to the reform. The closer the cohorts are the more difficult it is to find a significant effect. I therefore use adjacent cohort windows of 4 years and estimate equation (1) defining as the control group ($E_{ic} = 0$) individuals born between $t - 5$ and $t - 1$, and as the treatment group ($E_{ic} = 1$) individuals born between t and $t + 4$. Panel C reports the estimates of the interaction term for each year t. The only significant and large estimates are found for low family-background individuals when comparing the cohort windows 1972-1976 to 1967-1971 and 1973-1977 to 1968-1972. In these cases the treated group is composed of individuals who were on average 8 years old when the reform took place, while those in the control group were on average 13 years old. Once again, this suggests that the reform had a negative long-run effect on low family-background individuals, consistent with the results in Panels A and B.

Figure 5: The effect of the 1981 reform altering the treatment and control groups
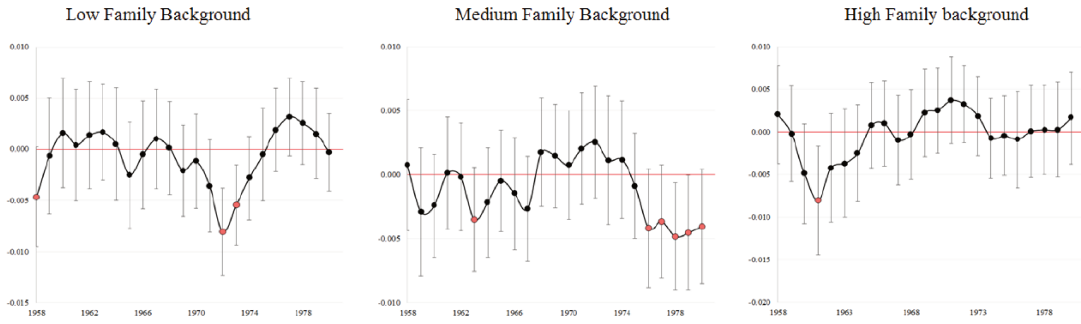
(a) Placebo: Reform in different years
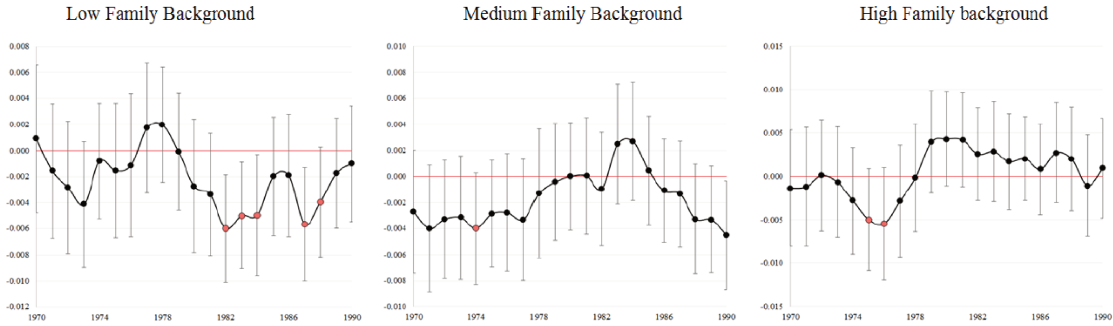
(b) Comparison group: Cohort 1950-1957
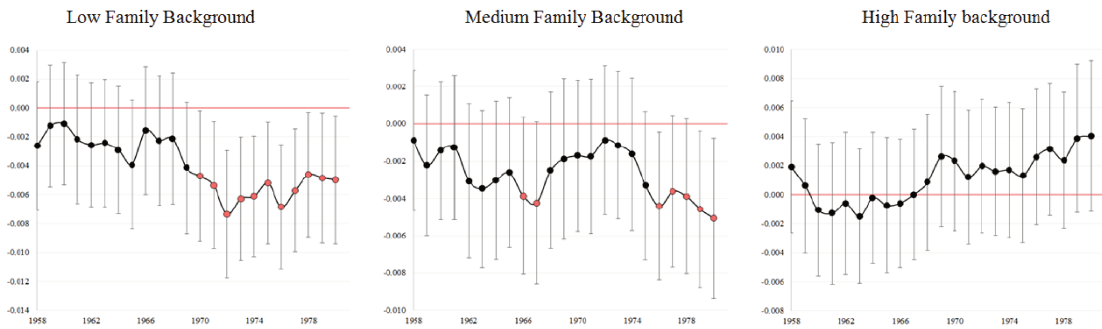
(c) Adjacent windows of 5-year cohorts

Note: Each point represents an estimated coefficient from a separate regression. Panel A shows the estimates of the interaction term from equation (1) (i.e., the effect of the reform on life satisfaction) as if the reform occurred in different years. For year t the control group (E=0) is composed of individuals born between t-20 and t-16 and the treatment group (E=1) of those born between t -10 and t-6. Panel B displays the estimates of the interaction term from equation (1) for different cohort windows setting as the reference group individuals born in 1950-1957. The estimate of year t is obtained by estimating equation (1) with a control group (E=0) composed of individuals born in 1950-1957 and a treatment group (E=1) of those born between t and t+4. Panel C displays estimates of the interaction term from equation (1) for different cohort windows that are contiguous. The estimate of year t is obtained by estimating equation (1) with a control group (E=0) composed of individuals born between t-5 and t-1 and a treatment group (E=1) of those born between t and t+4. There vertical bars are the 95 percent confidence intervals; red points indicate that the estimates are significant at least at the 10 percent level.

5.4 Heterogeneous effects

I look at potential heterogeneous effects depending by sex in columns 1 and 4 of Table 2 and by mobility patterns in columns 5 to 8. I consider individuals as internal migrants if they currently live a municipality other than the one in which they were born. Given that there is no information of the moment they moved, I assign them the rise of privatization (1091-1988) of their birth municipality. [17]

The impact of the reform is not significantly different for men and women of low family backgrounds. The coefficient for men displayed in column 2, although insignificant at conventional levels, presents a p-value of 0.11. The positive effect previously found on high family background individuals seems to be completely driven by men as shown in column 4. Regarding mobility patterns, the negative effect on individuals from low family backgrounds is explained by those who currently live in their municipality of birth. Individuals who moved, as shown in column 6, do not seem to have been affected by the reform in the long run. This may suggest that individuals could escape the pernicious effects of the reform by abandoning those areas, or alternatively, that individuals who select into migration have other traits that could compensate the negative impact of the reform. Among high family background individuals however the effects are not significantly different between movers and stayers. This may suggest that positive impact of the reform on this group is carried with them regardless of their location.

Lastly, I focus on the group individuals who left their municipality of birth and estimate equation (1) using the rise in private enrollment (1981-1988) in their current municipality of residence. The results appear in columns 9 and 10. Low family-background individuals who moved to high privatization areas are considerably worse-off than similar individuals (from the same municipalities of origin) who moved to low privatization areas. On average, a 10 percentage point increase in subsidized private enrollment is associated with a fall in life satisfaction of 11.1 percent of a standard deviation. These findings suggest that individuals make bad choices at the time of migrating and that they would have been better off staying in their municipalities of origin. On the contrary, high family-background individuals who moved seem unaffected by the degree of past privatization in the area where they live.

---

[17]This is the case for 45 percent of the sample.

Table 2: Effect of the 1981 voucher by sex and mobility patterns

	Life satisfaction (standardized)									
	Low FB		High FB		Low FB		High FB		Movers	
	Women	Men	Women	Men	Stayers	Movers	Stayers	Movers	Low FB	High FB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Exposed to the reform	-1.59	0.9344	-0.8205	2.4071	-0.8244	-0.6315	1.0726	-0.8646	0.3694	-0.8133
(born in 1972-1976 rather than 1962-1966)	(0.61)	(0.755)	(0.755)	(0.19)	(0.755)	(0.946)	(0.755)	(0.755)	(0.913)	(0.755)
Δ Private enrollment (%) * Exposed to the reform	-0.0053	-0.0079	0.0001	0.0081	-0.0093*	0.0000	0.0025	0.0040	-0.0111*	-0.0018
	(0.154)	(0.476)	(0.966)	(0.333)	(0.07)	(0.869)	(0.804)	(0.555)	(0.07)	(0.804)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	5640	2005	4507	2448	4435	3210	3140	3815	3210	3815
Adjusted R2	0.021	0.045	0.020	0.063	0.043	0.012	0.036	0.032	0.016	0.036

Note: All columns rely on the model in equation (1). The dependent variable is a standardized measure of life satisfaction. Stayers represent individuals who live in the same municipality they were born. Movers represent individuals live in a municipality other than that where they were born. Regardless of their mobility status, in columns (1) to (8) all individuals are assigned the increased in private enrollment (1981-1988) in their birth municipality. In columns (9) and (10) individuals are assigned the rise in private enrollment (1981-1988) of the municipality they currently live in. FB stands for family background based on parental education. Standard errors clustered at the municipality-decade level appear in parentheses. Standard errors clustered at the municipality-decade level and corresponding p-values adjusted for multiple hypothesis testing in parenthesis (following Anderson, 2008). Statistical significance: * p<0.1, ** p<0.05, *** p<0.01. Individual controls include a dummy for aboriginal origin, living with both parents during childhood, parental schooling, and age and age-squared. The municipality trends include the evolution of population density, share of children, urbanization, employment, average schooling and schooling inequality from 1970 to 1982.

5.5   Potential mechanisms

One way to understand through which channels the 1980s school privatization had an impact on individuals' well-being up to adulthood is to look at how key factors for well-being and may have been affected . Table 3 shows the results of estimating equation (1) for different adult outcomes such as health (on a 1-7 scale), years of education, household income, a dummy for being single (rather than in partnered), the number of children and a dummy for having migrated (i.e., leaving the municipality of birth). In the case of low family-background individuals, none of these factors were significantly affected by the reform: the interaction coefficient is always insignificant and close to zero. This is striking to the extent that these individuals appeared to be the most affected by the 1980s rise in private enrollment. The same picture arises for high family background individuals for whom, with the exception of household income, none of the interaction terms are

relevant.[18] A 10 percentage point rise in 1980s privatization is here associated with current household income that is 7.6 percent higher.

Table 3: Effect of the 1981 voucher reform on different adult outcomes

	Panel A. Low Family Background						
	Healh (1)	Schooling (2)	Hh. Income (3)	Single (4)	Children (5)	Urban (6)	Migrated (7)
Exposed to the reform	-0.4624***	-4.8111***	-1.0823***	-0.0295***	-0.3684***	-0.2823	-0.0939**
(born in 1972-1976 rather than 1962-1966)	(0.001)	(0.001)	(0.001)	(0.007)	(0.001)	(0.581)	(0.016)
Δ Private enrollment (%) * Esposed to the reform	0.0011	-0.0065	0.0003	0.0009	-0.0009	0.0011	-0.0003
	(0.901)	(0.901)	(0.901)	(0.63)	(0.901)	(0.901)	(0.901)
Number of observations	7593	7617	7616	7624	7624	7624	7624
Adjusted R2	0.050	0.218	0.125	0.007	0.147	0.161	0.209
	Panel B. High Family Background						
Exposed to the reform	-0.3156***	-1.1799***	-0.8090	-0.0835***	0.5172***	0.1455	0.0876
(born in 1972-1976 rather than 1962-1966)	(0.001)	(0.001)	(0.146)	(0.001)	(0.001)	(0.124)	(0.799)
Δ Private enrollment (%) * Esposed to the reform	0.0015	0.0145**	0.0076	-0.0010	-0.0026	0.0010	-0.0008
	(0.989)	(0.692)	(0.49)	(0.49)	(0.726)	(0.49)	(0.989)
Number of observations	6898	6912	6911	6922	6922	6922	6922
Adjusted R2	0.058	0.251	0.182	0.014	0.100	0.182	0.115

Note: Each column (in each Panel) uses the model in equation (1) for a different dependent variable. FB stands for family background based on parental education. Standard errors clustered at the municipality-decade level and corresponding p-values adjusted for multiple hypothesis testing in parenthesis (following Anderson, 2008). Statistical significance: * p<0.1, ** p<0.05, *** p<0.01. Individual controls include a dummy for female, aboriginal origin, living with both parents during childhood, parental schooling, and age and age-squared. The municipality trends include the evolution of population density, share of children, urbanization, employment, average schooling and schooling inequality from 1970 to 1982.

We can interpret these results in two ways. First, they provide additional support to the idea that there are no unobserved trends driving the results, as otherwise we could expect they would also affect other relevant outcomes such as health, schooling or income. The case of subjective health is particularly important given that it is an important driver of life satisfaction.[19] Notice that there do not seem to be general trends in these outcomes for the cohorts considered as the coefficient for the variable of "Exposed to the reform´´ tends to be insignificant. Second, the fact that in general

---

[18]Although in the case of years of schooling the interaction coefficient is significant, its small magnitude makes it irrelevant.

[19]The correlation between life satisfaction on health is 0.278 (p-value,0.001).

very few of the outcomes considered are affected by the reform poses the puzzle of understanding through which channels the impact on subjective well-being passes.

Some alternative mechanisms include assortative mating, income comparisons and views on inequalities. Since schools became more stratified as privatization increased, we may expect that students attending school in the 1980s interacted with peers of more similar backgrounds, at least compared to those attending school before the reform. Hence, as a result individuals may have end up with partners of more similar socio-economic status. One way to test this hypothesis is to look at the relative distance between couples' years of schooling, income or parents' education. For each of these outcomes I apply the model in equation (1) for low family-background individuals. The results show that there is no relationship between the 1980s privatization and couple similarity in these dimensions.[20]

Being conscious of segregation and perceiving it as a bad could make individuals more sensitive to social justice. For instance, they may be tempted to see other people's success as unfair and thus be more prone to a negative effect of income comparisons. Similarly, they could be more critical of high levels of income inequality or other forms of inequities. To test this hypothesis I construct a measure of comparison income26 based on people of similar age in each province and I calculate a measure of income inequality in each municipality.[21] I then introduce a triple interaction between these variables, the dummy for being exposed to the reform and the variable measuring the rise in subsidized private enrollment in the 1980s.28 The results show that none of the estimates of the triple interactions are significant, therefore rejecting the previous hypothesis. Note that neither the comparison-income variable nor income inequality have a direct relationship with life satisfaction in this context.

5.6 Robustness checks

Table 4 presents the results of estimating equation (1) adding sequentially all the variables, fixed effects, municipality controls and trends. The regressions focus on the sample of individuals born

---

[20]Looking at the absolute distance does not yield significant results either. The only significant result is for parental schooling, but the small size of the coefficient makes this irrelevant. The coefficient implies that increasing subsidized private enrolment up to 100 percent would reduce the gap between the couples' parents' schooling by 7 percent.

[21]For an individual i from cohort c the comparison income is the average income of all individuals in her province from cohorts comprised between $c - 4$ and $c + 4$. I calculate the Gini coefficient of individual income and household income per capita for each municipality.

in 1962-1966 and 1972-1976 with no missing information, either from low or high family backgrounds. The municipality controls here include measures of population, urbanization, religiosity, average schooling, schooling inequality,[22] schooling growth, mortality, fertility and political repression during the 1970s and 1980s. I also add the number of subsidized private schools in 2013 and their enrollment as a proxy for competition and current privatization. The municipality trends include measures of population density, share of children, urbanization, employment, share of indigenous, schooling and schooling inequality. Focusing on low family background individuals, even without any fixed effects or municipality controls, the interaction coefficient is negative and significant as presented in column 1. While municipality controls only increase the precision of the estimate marginally (column 5), replacing the controls by municipality fixed effects as in column 7 makes the interaction coefficient increase considerably in size and significance. The introduction of municipality trends in column 9 also raises the precision of the estimate. Regarding high family background individuals, while the coefficient is always non-negative, it only becomes large and marginally significant when including municipality fixed effects. As before, including municipality trends helps improving precision.

---

[22]For the first variable I take the average parental schooling of all individuals born in the 1970s. I also calculate the *gini* of parental schooling for these cohorts. Finally, I calculate the percentage increase in average parental schooling between individuals born in the 1950s and 1970s

Table 4: Robustness: Effect of the 1981 voucher reform on adult life satisfaction (2011-13)

	Life satisfaction (standardized)									
	Low FB	High FB	Low FB	High FB	Low FB	High FB	Low FB	High FB	Low FB	High FB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Exposed to the reform	0.0203	0.0432	-0.7931	-0.0273	-0.7740	-0.0522	-0.7765	0.3117	-0.4819	0.1990
(born in 1972-1976 rather than 1962-1966)	(0.0958)	(0.0841)	(0.6674)	(0.7408)	(0.6660)	(0.7371)	(0.6825)	(0.7717)	(0.6708)	(0.7657)
Δ Private enrollment (%)	-0.0005	-0.0038*	-0.0003	-0.0038*	-0.0039	-0.0032				
	(0.0022)	(0.0023)	(0.0022)	(0.0023)	(0.0028)	(0.0029)				
Δ Private enrollment (%) * Exposed to the reform	-0.0047*	0.0022	-0.0049*	0.0022	-0.0052*	0.0030	-0.0053***	0.0036*	-0.0059***	0.0038*
	(0.0027)	(0.0030)	(0.0027)	(0.0030)	(0.0027)	(0.0027)	(0.0017)	(0.0020)	(0.0018)	(0.0021)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed effects	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Municipality Controls	No	No	No	No	Yes	Yes	No	No	No	No
Municipality Trends	No	No	No	No	No	No	No	No	Yes	Yes
Number of observations	7624	6922	7624	6922	7624	6922	7624	6922	7624	6922
Adjusted R2	0.015	0.024	0.015	0.023	0.029	0.032	0.019	0.027	0.028	0.032

Note: These are different variants of equation (1). The dependent variable is a standardized measure of life satisfaction. FB stands for family background based on parental education. Standard errors clustered at the municipality-decade level are in parentheses. Statistical significance: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Individual controls include a dummy for female, aboriginal origin, living with both parents during childhood, parental schooling, and age and age-squared. The municipality-level controls include the total number of students in 1981, the share of urban students in 1981, the increase in subsidized private enrollment between 1981 and 1988, the share of students attending subsidized private schools in 1981 and 2012, the number of subsidized private schools in 1991 and 2012, average schooling in the 1970s, schooling inequality in the 1970s, and the share of indigenous population in the 1970s. The provincial-level controls are total population in 1970, the share of non-religious people, the support for Allende (left) in the 1970 presidential election, the number of missing and assassinated people by the regime during the dictatorship, and mortality and fertility rates. The municipality trends include the evolution of population density, share of children, urbanization, employment, average schooling and schooling inequality from 1970 to 1982. All data sources are detailed in Table A.1 in the Appendix.

6 Conclusion

This paper has analyzed the impact of the 1981 education reform in Chile on long-term happiness. Exploiting geographical variation in the penetration of subsidized private schools during the 1980s and comparing cohorts that were differently exposed depending on their year of birth, I find suggestive evidence that the reform had a long-lasting negative effect on life satisfaction for individuals of low family background. On average, a 10 percentage point increase in subsidized private enrollment in their municipality when they were young is associated with a fall in adult life satisfaction in 5.9 percent of a standard deviation. This impact is equivalent to the effect of lower household income of around 30 percent or reducing education by 3 years. On the contrary, high family-background individuals may have benefited from the reform, although the evidence is less robust. The impact differs markedly by sex and migration. The negative effect on those with low family background is entirely driven by those who stayed in their municipality of birth. The positive effect for high family-background individuals is only found for men and individuals who migrated to another municipality. Focusing on those who moved, those from low family backgrounds ending up in high-privatization areas are substantially worse-off. In this case, a 10 percentage point increase in private enrollment is associated with a reduction in adult life satisfaction of 11.1 percent of a standard deviation. Changing the cohort windows defining the treated and control groups as if the reform had occurred earlier or later produces results that are consistent with the main findings.

The original objective of the 1981 educational reform, according to its proponents, was to make the educational system more efficient and raise educational outcomes. While total expenditure fell considerably, there was no meaningful effect on performance or achievement. The explanation is that private schools competed on the basis of students' socio-economic status rather than on quality. We can draw important policy implications from the evidence presented here. First, when analyzing educational policies, it is not sufficient to look at educational outcomes: students can be affected in many other dimensions that ultimately impact subjective well-being. Second, it is not sufficient either to look only at contemporaneous outcomes, as the impact of a policy change can manifest itself many years later, as with the voucher system. For instance, certain policies could lead students to choose some careers over others and earnings differentials may only appear many years afterwards. Furthermore, when there exist permanent impacts throughout life,

the potential welfare gains or losses can be enormous and need to be taken into account. Finally, under an appropriate setting in which information is well collected, subjective well-being data can be a useful way of taking multiple dimensions into account and examining the long-term welfare consequences of educational policies.

The findings of this paper highlight how a policy meant to benefit the poorest students by providing them with the opportunity to attend better schools had pernicious effects due to bad design. Choice was never an option for the least well-off and, as result, they did not benefit from the reform. The fact that this policy change had a permanent effect on adult life satisfaction calls for a well-thought design of educational policies: otherwise, even well-indention measures may compromise individual well-being throughout their lives. Future research should focus on better exploring and understanding the mechanisms at play, and the role played by family background in the outcomes.

References

Anderson, M.L., 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects", Journal of the American statistical Association, 103(484), pp.1481-1495.

Auguste, S., and J. P. Valenzuela, 2004. "Do students benefit from school competition? The Chilean experience´´, PhD dissertation, University of Michigan: Ann Arbor, MI.

Bellei, C., 2009. "Expansion of private schools and educational improvement in Chile. An evidence-based evaluation´´, Revista pensamiento educativo, 40(1), 1-37.

Böhlmark, A., Holmlund, H. and M. Lindahl, 2016. "Parental choice, neighbourhood segregation or cream skimming? An analysis of school segregation after a generalized choice reform´´, Journal of Population Economics, 29(4), 1155–90.

Bravo, D., Contreras, D. and C. Sanhueza, 1999. "Rendimiento educacional, desigualdad y brecha de desempeño público/privado: Chile 1982-1997´´, Departamento de Economía, Universidad de Chile.

Brum, M., 2018. "Do dictatorships affect people's long term beliefs and preferences? An empirical assessment of the Latin American case´´, Serie Documentos de Trabajo; 18/18.

Carnoy, M., 1998. "National Voucher Plans in Chile and Sweden: Did Privatization Reforms Make for Better Education?´´, Comparative Education Review, 42(3), 309-337.

Carnoy, M. and P. McEwan, 2000. "The effectiveness and efficiency of private schools in Chile's voucher system´´, Educational Evaluation and Policy Analysis, 22, 213– 239.

Chetty, R., Hendren, N. and L.F. Katz, 2015. "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment´´, Technical report, National Bureau of Economic Research.

Chumacero, R., Gomez, D. and R. Paredes, 2011. ″I would walk 500 miles (if it paid): Vouchers and school choice in Chile´´, Economics of Education Review, 30, 1103– 1114.

Contreras, D., Sepulveda, P. and S. Bustos, 2010. ″When Schools Are the Ones that Choose: The Effects of Screening in Chile´´, Social Science Quarterly, 91(5), 1349 – 1368.

Cox, C., 2003. ″Las políticas educacionales de Chile en las últimas dos décadas del siglo XX´´, Políticas educacionales en el cambio de siglo. La reforma del sistema escolar en Chile, 19-113.

Echenique, F., Fryer, R. G. and A. Kaufman, 2006. ″Is school segregation good or bad?´´, American Economic Review, 96(2), 265–269.

Elacqua, G., 2012. ″The impact of school choice and public policy on segregation: Evidence from Chile´´, International Journal of Educational Development, 32(3), 444-453.

Espinola, V., 1993. ″The educational reform of the military regime in Chile: The school system's response to competition, choice, and market relations´´, Ph.D. dissertation, University of Wales College Cardiff.

Gauri, V., 1998. ″School Choice in Chile´´, University of Pittsburgh Press.

González, F., Prem, M. and Urzúa I, F., 2020. ″The privatization origins of political corporations´´, Journal of Economic History, 2020, 80(2), 417-456..

Guryan, J., 2004. ″Desegregation and Black Dropout Rates´´, American Economic Review, 94(4), 919-943.

Hsieh, C-T. and M. Urquiola, 2006. ″The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program´´, Journal of Public Economics, 90, 1477–1503.

Johnston, R., Wilson, D. and Burgess, S., 2007. "Ethnic segregation and educational performance at secondary school in Bradford and Leicester´´, Environment and Planning, 39(3), 609–629.

Kling, J. R., Ludwig, J. and L. F. Katz, 2005. "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment´´, Quarterly Journal of Economics, 120(1), 87-130.

Ludwig, J., Greg J. D., Gennetian, L.A, Katz, L. F., Kessler, R. C., Kling, J. R. and Sanbonmatsu. L. 2012. "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults´´, Science, 337(6101), 1505–10.

McEwan, P., Urquiola, M. and E. Vegas, 2008. "School choice, stratification, and information on school performance: Lessons from Chile´´, Economia, 8(2), 1-27.

Mediavilla, M. and A. Zancajo, 2015. "Is there real freedom of school choice? An analysis from Chile´´, Universidad de Barcelona WP.

Mizala, A. and P. Romaguera, 2000. "School performance and choice: The Chilean experience´´, Journal of Human Resources, 35(2), 392-417.

Portales, J. and J. V. Heilig, 2015."Understanding Universal Vouchers and Urban Public Schools in Santiago de Chile: Educational Administrators' Responses to Choice´´, REMIE Multidisciplinary Journal of Educational Research, 5(2), 196–239.

Prawda, J., 1993. "Educational decentralization in Latin America: lessons learned.´´ International Journal of Educational Development, 13(3), 253-264.

Reber, S. J., 2010. "School desegregation and educational attainment for blacks´´, Journal of Human Resources, 45(4), 893–914.

Sanbonmatsu, L. et al., 2012. "The long-term effects of moving to opportunity on adult health and economic self-sufficiency´´, Cityscape, 109-136.

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

Söderström, M. and R. Uusitalo, 2005. "School choice and segregation: evidence from an admission reform´´, Technical report, Working Paper, IFAU-Institute for Labour Market Policy Evaluation.

Torche, F., 2005. "Privatization reform and inequality of educational opportunity: The case of Chile´´, Sociology of Education, 78(4), 316–343.

Valenzuela, J.P., Bellei, C. and D. De los Rios, 2014. "Socioeconomic school segregation in a market-oriented educational system. The case of Chile´´, Journal of Education Policy, 29(2), 217-41.

Weiner, D. D., Lutz, B. F. and J. Ludwig, 2009. "The Effects of School Desegregation on Crime´´, Technical report, National Bureau of Economic Research.
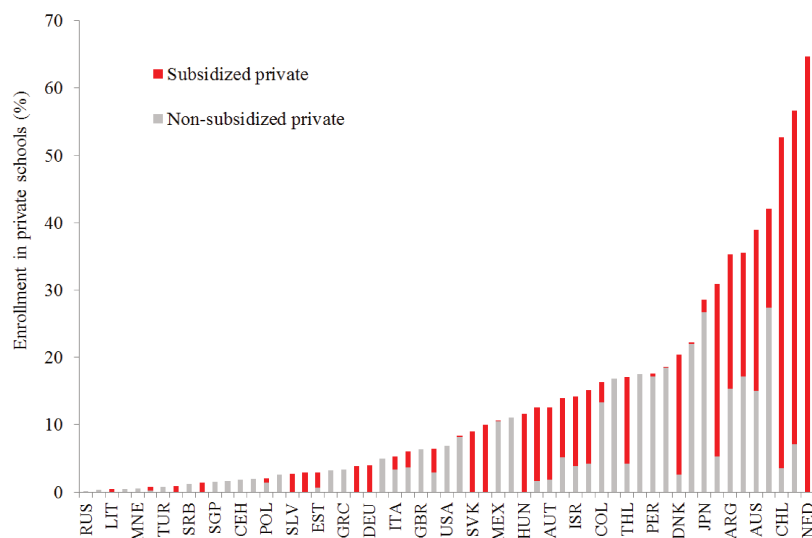
A Online Appendix

Additional Figures

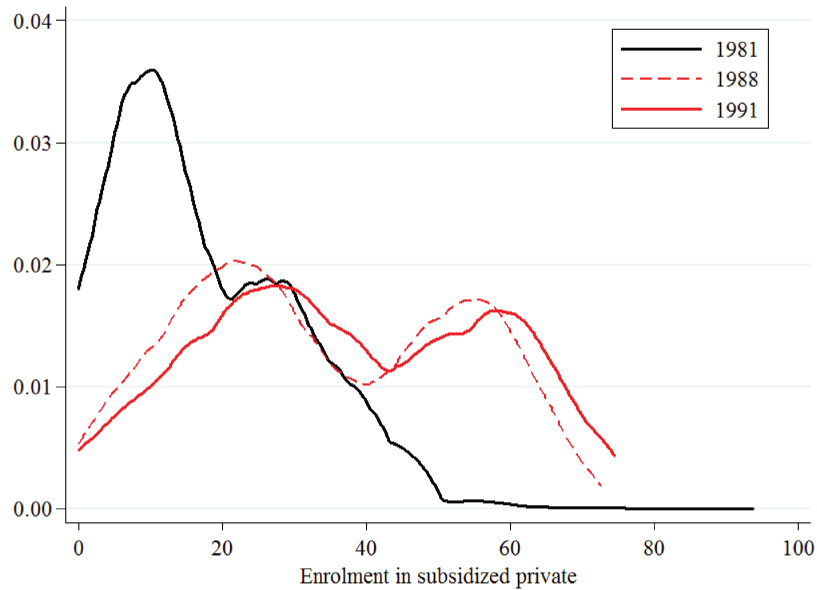Figure A.1: School segregation across countries, 2012
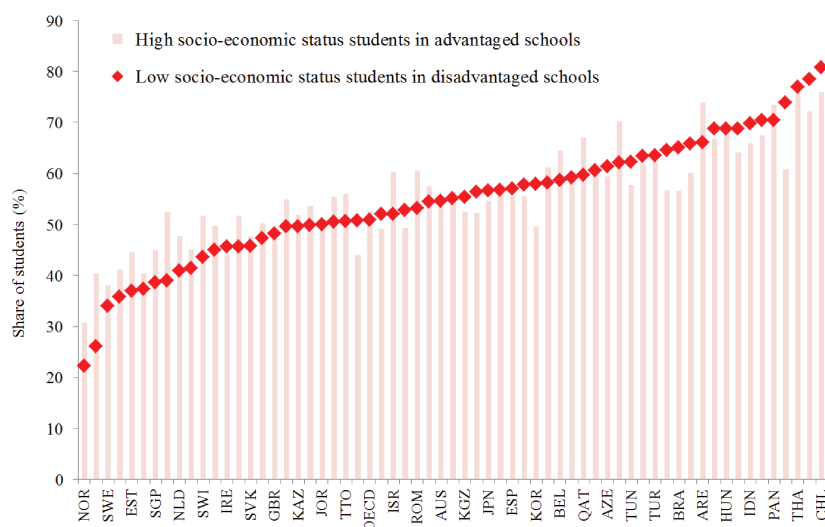
Note: Own elaboration using PISA 2012 data. Return to page 1.

Figure A.2: Enrollment in private schools across OECD countries, 2014
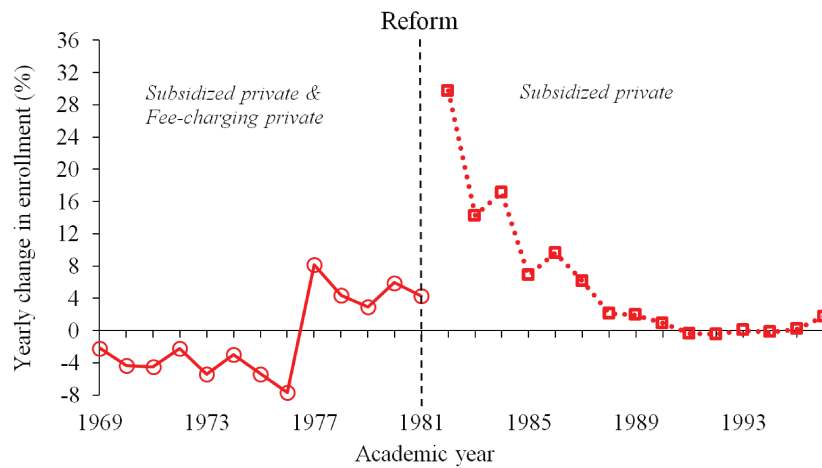

Note: Own elaboration using OECD data. Return to page 1.

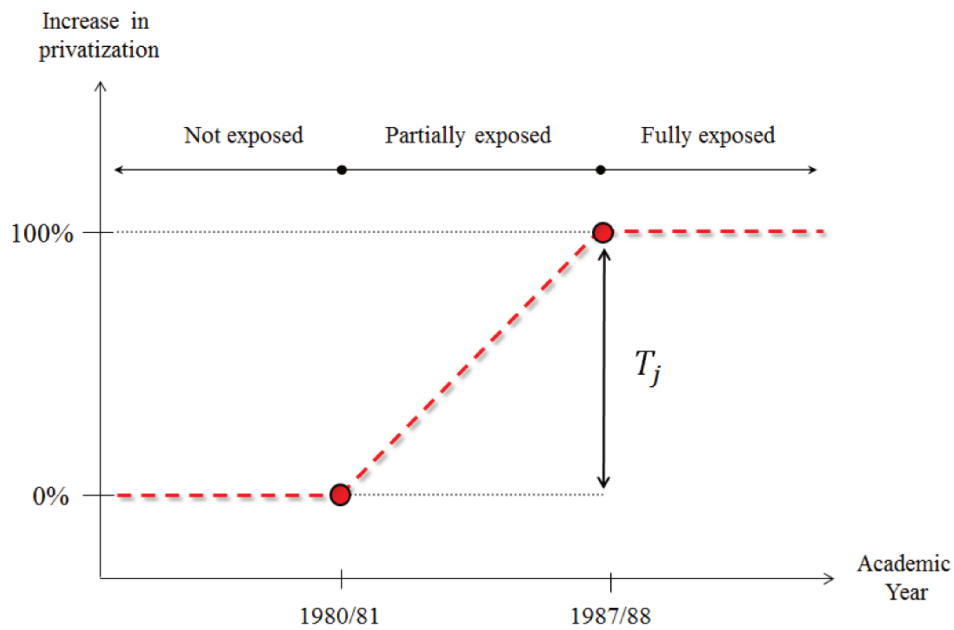Figure A.3: Distribution of enrollment in subsidized private schools

Note: Data from Ministry of Education. Return to page 6.

Figure A.4: Yearly increase in subsidized private enrollment

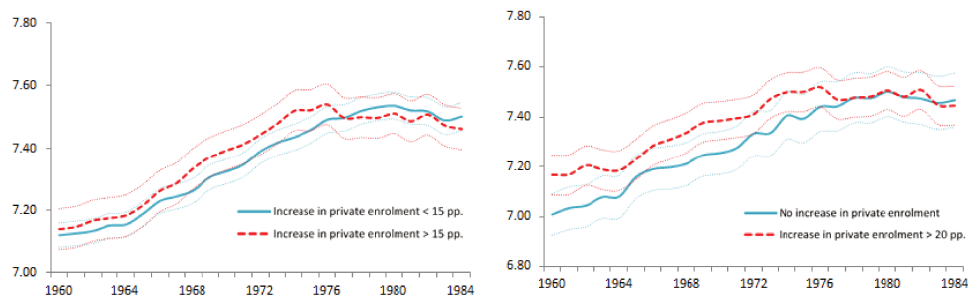*Note:* Data from Ministry of Education. Return to page 6.

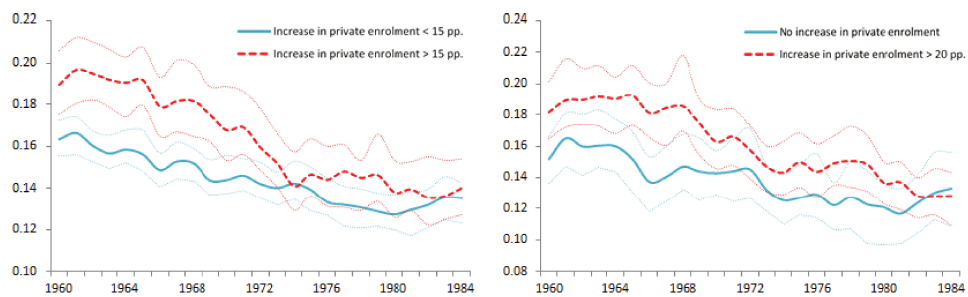Figure A.5: Exposure to rising segregation due to the 1981 reform

Figure A.6: Adult life satisfaction by birth cohort

Table 1: Summary of data sources

Data source	Information	Unit of observation	Years
National Household Survey (CASEN)	Life satisfaction	Individual	2011, 2013
	Demographics	Individual	2009,2011,2013, 2015
	Parents' schooling	Individual	2009,2011,2013, 2015
	Income and others	Individual	2009,2011,2013, 2015
Ministry of Education	Matriculation records	County	1981, 1988, 1991, 2012
	School records	County	1990, 2012
National Statistics	Mortality and fertility	Province	1980
Bank of Chile	GDP growth	Region	1985-1990
Census	Population	Province	1970, 1982, 1992, 2002
	Urbanization	Province	1970, 1982, 1992, 2002
	Religiosity	Province	1970, 1982, 1992, 2003
	Unemployment rate	Province	1982, 1992, 2002
Electoral Service	Electoral results	Province	1970
Museum of Memory and Human Rights	Political deaths and missing	Province	1973-1990

Note: Return to page .

Table 2: Summary statistics of individual characteristics

Variable	Obs	Mean	Std. Dev.	Min	Max	Low FB	Medium FB	High FB
Life satisfaction	58824	7.36	2.13	1	10	7.10	7.32	7.70
Women	58824	0.68	0.47	0	1	0.72	0.68	0.64
Indigenous	58824	0.13	0.33	0	1	0.17	0.12	0.08
Parents' schooling	58824	7.19	4.62	0	22	2.76	6.87	12.64
Age	58824	40.05	6.94	27	52	39.70	40.53	39.79
Birth comuna	58824	0.55	0.50	0	1	0.59	0.57	0.47
Birth year	58824	1972	7	1961	1984	1972	1971	1972
Schooling	58754	11.20	3.68	0	22	9.49	11.00	13.41
Household income per member	58824	267195	437942	0	46100000	177112	232069	416181
Single	58824	0.22	0.41	0	1	0.23	0.21	0.21
Number of children	58824	1.32	1.11	0	11	1.41	1.30	1.24
Urban area	58824	0.80	0.40	0	1	0.73	0.79	0.91
Health status	58602	5.58	1.23	1	7	5.49	5.55	5.73

Note: Household survey CASEN of 2011 and 2013. Note: FB stands for family background. Household income per member is expressed in current Chilean pesos. Life satisfaction and health status are self-reported..

Table 3: Summary statistics of enrollment data

Variable	County level					Individual level		
	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.
Enrolment in public schools, 1981	346	86.83	15.81	6.25	100.00	58824	82.12	14.58
Enrolment in subsidized private schools, 1981	346	10.86	13.87	0.00	93.75	58824	13.10	12.51
Enrolment in traditional private schools, 1981	346	2.31	7.23	0.00	61.05	58824	4.79	7.64
Enrolment in public schools, 1988	346	78.31	20.91	24.71	100.00	58824	69.36	20.14
Enrolment in subsidized private schools, 1988	346	18.78	18.26	0.00	72.72	58824	24.59	17.02
Enrolment in traditional private schools, 1988	346	2.17	7.77	0.00	63.51	58824	4.52	8.03
Change in subsidized private enrolment, 1981-1988	346	7.92	11.78	-93.75	54.16	58824	11.49	9.33

Note: Enrollment records from the Ministry of Education of Chile. Note: the columns under "Individual level" show the statistics once the municipality level data has been assigned to the individuals in the sample..

Table 4: Effect of the 1981 voucher reform using different measures of exposure

	Life satisfaction (standardized)					
	Low FB	High FB	Low FB	High FB	Low FB	High FB
	(1)	(2)	(3)	(4)	(5)	(6)
Years of exposure	0.205	-0.1246	0.0946	-0.0175	0.1538	-0.0922
	(0.2214)	(0.2518)	(0.0590)	(0.0668)	(0.1606)	(0.1827)
Δ Private enrollment (%) * Years of exposure	-0.0007***	0.0003	-0.0042**	-0.0008	-0.0011***	0.0004
	(0.0002)	(0.0002)	(0.0021)	(0.0016)	(0.0003)	(0.0003)
Years of exposure:	Fully exposed if in school post-1981		Fully exposed if in school post-1988		Fully exposed if in school post-1988 Partially exposed if in school in 1981-88	
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Municipality fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Municipality trends	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R2	0.027	0.025	0.027	0.025	0.027	0.025
Number of observations	11414	10415	11414	10415	11414	10415

Note: Columns 1 to 6 estimate the model in equation (1) with alternative measures of exposure to the reform. The dependent variable is a standardized measure of life satisfaction. FB stands for family background based on parents' education. The sample comprises all individuals born between 1962 and 1978. Standard errors clustered at the municipality-decade level in parenthesis. Statistical significance: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Individual controls include a dummy for being women, pertaining to the aboriginal populations, living with both parents during childhood, parental schooling, age and age squared.

Summary:
Essays in Applied Economics

This thesis exploits quasi-natural experiments, both historical and contemporary, to provide causal evidence on broad questions at the frontier of the development and political economy literature. In particular, it delves into the long-run effects of emigration, the formation of identities and political preferences, and the long-term impacts of school segregation. The first chapter examines the consequences of emigration for human capital accumulation at origin, taking as a laboratory the Galician diaspora formed at the beginning of the XX century. Leveraging data spanning over one hundred years, it provides the first evidence of a positive impact of emigration in the communities of origin across several generations. The second chapter investigates whether the emotional trigger of sports can influence identity and political preferences. Focusing on Catalonia and Football Club Barcelona (FCB) as an ideal setting, I show empirically that FCB performance is directly linked with feelings of Catalan identity and support for secession. Furthermore, I explore the mechanisms behind this connection and its potential implications for electoral results. The third chapter takes advantage of a major reform under the Pinochet's dictatorship, which introduced a system of school vouchers, to analyze the long-term effects of school segregation on subjective well-being. The findings show that the reform had unintended consequences, hurting children from poorer family backgrounds who today, more than 30 years later, have significantly lower levels of life satisfaction.

Résumé: Essais en Économie Appliquée

Cette thèse utilise des expériences quasi-naturelles, historiques et contemporaines, pour aborder de manière causale grandes questions à la frontière des champs du développement et l'économie politique. En particulier, elle se penche sur les effets à long terme de l'émigration, la formation des identités et des préférences politiques, et les impacts à long terme de la ségrégation scolaire. Le premier chapitre examine les conséquences de l'émigration sur l'accumulation du capital humain, en prenant comme laboratoire la diaspora galicienne formée au début du XXe siècle. En s'appuyant sur des données couvrant plus de cent ans, il fournit les premières preuves d'un impact positif de l'émigration sur les communautés d'origine sur plusieurs générations. Le deuxième chapitre examine si les shocks émotionnelles liés aux sports peuvent influencer l'identité et les préférences politiques. En me concentrant sur la Catalogne et le Football Club Barcelona (FCB) comme contexte, je montre empiriquement que les performances du FCB sont directement liées aux sentiments d'identité catalane et au soutien à la sécession. En outre, j'explore les mécanismes qui sous-tendent ce lien et ses possibles implications sur les résultats électoraux. Le troisième chapitre profite d'une grande réforme pendant la dictature de Pinochet, qui a introduit un système des chèques scolaires, pour analyser les effets à long terme de la ségrégation scolaire sur le bien-être subjectif. Les résultats montrent que la réforme a eu des conséquences involontaires, nuisant les enfants issus de familles pauvres qui affichent aujourd'hui, plus de 30 ans plus tard, des niveaux de bien-être subjectif nettement inférieurs.