PhD-FDEF-2024-022
The Faculty of Law, Economics and Finance

DISSERTATION

Defence held on 26 November 2024 in Belval, Luxembourg
to obtain the degree of

Docteur de l'Université du Luxembourg en Sciences Economiques

by

Etienne Bacher

Born on 8 June 1998 in La Rochelle (France)

Cross-border labor mobility, attitudes and political preferences

Dissertation defence committee

Dr. Michel TENIKUE, Dissertation supervisor
LISER

Prof. Michel BEINE, member of the thesis committee
Professor, University of Luxembourg

Prof. Hillel RAPOPORT, member of the thesis committee
Professor, Paris School of Economics

Dr. Sandra SEQUEIRA
Associate Professor, London School of Economics

Prof. Skerdilajda ZANAJ
Professor, University of Luxembourg

Acknowledgements

I'll start with the people without whom this thesis wouldn't have existed at all, the three members of my thesis committee: Michel T., Michel B., and Hillel. I have learned a lot in this period and you have been essential to this. In particular, the first chapter of this thesis is joint work with Michel B. and Hillel and has been plagued by mistakes. Those were countless and painful every time I had to admit them, but they also forced me to improve my work organization and habits. I can only thank you both for the patience you have showed on this chapter and for all the advice and fruitful discussions that came from this.

I'd like to thank the two other members of the defence committee, Sandra Sequeira and Skerdi Zanaj, for the thoughtful comments and ideas that you have shared with me.

Sandra also had the kindness of sponsoring my research visit at the London School of Economics, for which I'm extremely grateful. I met brilliant people there, and I particularly want to thank Eric and Marco for their time spent with me, Monika for helping me going through all the administrative process, and Gijs, Jorich, Parth, and Mina for sharing their office. London was also the place where I met an amazing group that made my stay much more enjoyable, the Dulwich Park Runners. Thank you David, and all the others, for making me feel welcome in this group. Those moments spent chatting and jogging under the British rain are some of my best memories of my stay.

Back to Luxembourg now, I want to deeply thank the people I've seen almost every day and with whom I shared a lot of feelings, from my hopes of publication in the first year to my despair about the state of academic research and of course my doubts about my capacity to produce good research. When this was the case, FIFA, chess, and table tennis came to my rescue. So Felix S., Julio, Carlotta, Narcisse, Lucas, Felix G., Maggie, Alex, Jules, Juliette, David, Jongoh, Salvatore, thank you! Thank you Martin, not only for being a friend and a great coauthor, but also for making me a better researcher and for remotivating and encouraging me a countless number of times.

Of course, I'm not forgetting those at Kirchberg, so close (let's be honest, Luxembourg is not that big) and yet so far. Thank you Rian, Adrij, Aleksa, Ariane, Silvia, Andreas, and Rana!

This is a good occasion to thank Frédéric and Michel B. for setting up the ACROSS program. They have provided help, encouragement, and above all they have created a friendly environment composed of amazing people. Thank you to the admin teams at the university and at LISER, and particularly to Noémie and Anna.

There are also many people I haven't seen regularly, but with whom I had a great time every time we met, so thank you Laura, Anthony, Giorgia, Adrian, Gaspare, Giuseppe, Karan, Natella, and Axelle!

Outside work, I have had the chance of having great roommates who made my life much better: Lyacine, and in particular Paulo and Raoul, with whom the memory of listening Kompromat at max power will always stay.

From before Luxembourg, I first want to thank Simone Bertoli for being my main inspiration and motivation for doing research on migration and for helping me going to LISER. Thank you Adrien, Richard, Charlotte, Coline, Anaïs, Jules, Ambre, Alexis, Juliette, Eva, Daniel, Camille, Eugénie, Lucas, Muriel. We didn't see each other a lot, but every time we met was one of the best times in the year, despite my "research sucks" mood sometimes. Thank you also to Florent Bresson, who made me discover R and a new passion in my life.

And now's the time to thank all of those who don't know they have helped me finishing this thesis, and yet have spent most of my worktime with me. Some of them have motivated me many times when I was procrastinating: Two Steps from Hell, Piano Deuss, Hans Zimmer, Powerwolf, Patrik Pietschmann, Valentina Lisitsa, and many others. In the extremely unlikely event you end up reading this section (but isn't it extremely unlikely for anyone to read the acknowledgements of a PhD thesis?), I want you to know you were essential to this. Among this "immaterial" help I have received, an important part was from the R community, and in particular Vincent Arel-Bundock, Laurent Bergé, Grant McDermott, Will Landau, and the easystats team for building so many awesome tools for free (without the acknowledgement they deserve). Lastly, I deeply thank Alexandra Elbakyan for all her efforts in promoting and effectively making science more open.

Finally, I want to thank my family for supporting me from the beginning of this adventure. Papa, Maman, Hugues, Benoît, Clément, Matthieu, merci pour votre soutien, merci pour vos photos des Alpes, du Canada, du Laos, de voyage à vélo, merci pour ces randos, ces parties d'échecs, ces blagues nulles (même si je me défends pas mal dans ce domaine), merci pour tout.

Contents

List of Figures

List of Tables

Abstract

This dissertation consists of three chapters. Each chapter constitutes a self-contained study, tackling a salient subject within the economics of international migration.

⋄ **Chapter 1: Estimating the impact of xenophobia on immigration: evidence from terrorist attacks.** Co-authored with Michel Beine (University of Luxembourg) and Hillel Rapoport (Paris School of Economics).

This paper analyses the sensitivity of potential migrants to anti-immigration attitudes in destination countries. We use migration plan data from the Gallup World Polls and measure anti-immigration attitudes from the Eurobarometer. To overcome important endogeneity issues, we instrument attitudes towards immigration with the number of citizens of a European country killed in terrorists attacks outside Europe. We find that terrorist attacks tend to increase negative attitudes with respect to immigration in the origin country of the victims, which validates the strength of our instrument. Those attitudes then decrease the country's attractiveness for potential immigrants, no matter their skill level.

⋄ **Chapter 2: Populist leaders and international migration.**

While immigration is often considered as an important cause of populism, the consequences of populist leaders on international migration have received much less attention. In this paper, I explore the dynamic effects of populist leaders leaders on both immigration and emigration, how they affect the selectivity of immigration, as well as the underlying mechanisms. I use data on populist leaders and migration flows since 1960 and the generalized synthetic control method for identification. I find that having a populist leader significantly decreases the growth rate of immigration, in particular for low-skilled immigration. However, there is no effect on emigration growth rates. I use rich datasets on migration policies and migration intentions to explore potential mechanisms, and I find that right-wing populist

leaders implement more restrictive migration policies. In addition to reducing inflows, those policies also encourage low-skilled migrants who were already settled in the country to leave and go back to their home country. On the other hand, having a populist leader doesn't change the attractiveness of the country in the eyes of potential migrants.

⋄ **Chapter 3: Migrant Voices.** Co-authored with Martín Fernández (LISER).

Migrants can lead to profound economic and social changes in their home communities through the information they share with family and friends left behind. This paper leverages unique historical data and artificial intelligence to analyze the content of migrants' communications, the factors shaping them, and their potential effects at origin. We construct a novel dataset of over 6,000 letters from Irish emigrants in North America in the 19th and early 20th centuries. We characterize letter writers and their local environment using data from passenger lists and population censuses at destination and origin. Relying on Large Language Models (LLMs), we identify and classify topics in the letters. We first provide some novel descriptive evidence about migrants' communications. We then explore the data and find that some individual and local characteristics (e.g., gender, religion, the size of the Irish diaspora, and average incomes) significantly affect the presence and salience of topics related to economics, religion, and politics. We plan on expanding the analysis to explore how a large institutional change at destination (i.e., the expansion of Catholic churches across the US) may have affected social remittances and, in turn, religious outcomes in Ireland.

Declaration of co-authorship

This section contains an overview on the role of the co-authors in the realization of this thesis.

⋄ **Chapter 1: Estimating the impact of xenophobia on immigration: evidence from terrorist attacks.** Co-authored with Michel Beine (University of Luxembourg) and Hillel Rapoport (Paris School of Economics).

We had a similar idea around the same time and decided to explore it together. I handled the data manipulation and estimation side, with a close guidance of the co-authors. All technical aspects of the paper were discussed and jointly decided by the three co-authors. The draft was jointly written.

⋄ **Chapter 2: Populist leaders and international migration.**

This is a single-authored paper.

⋄ **Chapter 3: Migrant Voices.** Co-authored with Martín Fernández (LISER).

Martín had the idea to use Irish migrants' letters as the main context of the paper. All subsequent ideas and analysis are joint work. We both contributed to all parts of the research process, including the redaction of the draft.

Introduction

International migration has become increasingly important in the public debate in the last two decades. Following this trend, the scientific literature exploring the political effects of migration, in particular on voting behavior, and the determinants of attitudes towards immigration has grown dramatically.

The determinants of attitudes towards immigration

The question of whether attitudes towards migrants are caused by economic or cultural factors received a lot of attention. Economically, the main rationale is that migrants constitute an increase of the labor force and therefore lead to more competition on the labor market. Individuals should therefore oppose immigration of the skill group that threatens their employment: high-skilled labor should oppose high-skilled immigration and favor low-skilled immigration, and vice-versa (Mayda, 2006). The effect of immigration on the welfare state should also affect attitudes towards migrants, but in the opposite direction: high-skilled individuals, who contribute more to the welfare state, should be against the arrival of low-skilled migrants since they would need to pay more to support them. On the other hand, they should support the arrival of high-skilled migrants, who would contribute more to the welfare state than they benefit from it (Facchini and Mayda, 2008; Facchini and Mayda, 2012). Those interactions between the labour market effects and the welfare effects were first thought to be the main determinants of attitudes towards immigration. To those economic factors at the individual-level, we should also add those at the country-level: anti-immigration attitudes are usually amplified in times of economic hardship (Hainmueller and Hopkins, 2014).

Still, while the economic causes were the usual suspects, the cultural aspects started to be considered as the main determinant of anti-immigration attitudes (Hainmueller and Hangartner, 2013). The idea that migrants cannot or don't want to integrate culturally have become more prominent in the political discourse, especially regarding immigration from poorer and culturally distant countries. A seminal paper on the question of economic versus cultural determinants is Hainmueller and Hopkins (2015). In this paper, they implement a large-scale survey in the United States in which respondents compare pairs of migrant profiles containing economic and cultural information and decide which one they prefer. One of the main conclusions of this paper is that all respondents prefer high-skilled migrants to low-skilled ones, which goes against the hypothesis that natives base their attitude on their self-economic interest. However, it is worth noting that economic factors still play a non-negligible role (Alesina and Tabellini, 2024).

Following this evolution in the search for the determinants of attitudes towards migration, an increasing number of studies explored the question of whether contact between migrants and natives reduces or increases prejudice. According to the *contact hypothesis* (Allport, 1954), under some conditions, an increased contact between natives and migrants in everyday life should lead to a closer relationship between those groups, and therefore should make opinions towards migration more favorable. However, this increased contact could also exacerbate tensions between the two groups, a mechanism that is described in the *group threat theory* (Blalock, 1967). So far, the contact hypothesis has found some support in the literature (Steinmayr, 2021; Bursztyn et al., 2024) but it is not applicable in all places, even within a country (Dustmann et al., 2019).

The effect of attitudes on future migrants

The existing literature has mostly focused on the political effects of immigration, generally looking at the voting behavior of natives. It has also tried to determine the determinants of anti-immigration attitudes. Yet, the effect of those attitudes on future migration flows has received less attention. One of the goals of this dissertation is to contribute to filling this gap.

Chapters 1 and 2 focus on the effect of those anti-immigration attitudes on migration flows and migration plans. One of the main challenges when tackling this issue is disentangling the effect of those attitudes from the one of other factors, such as economic conditions. Each of those chapters uses a different different strategy.

In Chapter 1, we instrument attitudes towards immigration by the number of victims from terrorist attacks happening outside Europe. We first show that this is a strong and plausibly exogenous instrument. Then, we use the instrumented measures of attitudes towards immigration in Europe to analyze the effect on migration plans. We find that higher hostility towards migrants in a European country discourages migrants from going to this country. Interestingly, this effect is present no matter the skill level of potential migrants.

In Chapter 2, I use the synthetic control method (SCM) and several of its variants to overcome the endogeneity issue mentioned above. In this chapter, the focus is no longer on attitudes towards migrants but on having a populist leader. Using the transition from having a non-populist leader to having a populist leader as an event, I apply the SCM to build a synthetic counterfactual that mimics the treated country in terms of past immigration flows and economic situation. Applying this event-study analysis to a dozen of countries, I find that having a right-wing populist leader leads to a decrease in immigration. This effect is mostly driven by the fact that right-wing leaders implement stricter immigration policies.

Chapter 3 addresses the question of international migration and political attitudes and preferences from another point of view. In this chapter, we use historical data to explore the transmission of preferences, norms, and values between migrants and their home country. We use a novel dataset composed of over 6,000 letters written by Irish migrants living in North America between 1840 and 1930 to explore the content of migrants' communication. After validating that our data is representative from multiple ways, we analyze the importance of multiple topics in migrants' letters. In particular, we explore how personal characteristics as well as the socio-economic environment in which migrants live affect the content of their letters. We find that Irish migrants talk mostly about economics and religion, and that there exists substantial heterogeneity depending on personal characteristics, such as gender, and regional factors, such as the size of the Irish diaspora or the size of the income gap between migrants and natives.

References

Alesina, A. and M. Tabellini (2024). "The Political Effects of Immigration: Culture or Economics?" In: *Journal of Economic Literature* 62(1), pp. 5–46.

Allport, G. (1954). *The Nature of Prejudice*.

Blalock, H. M. (1967). *Toward a Theory of Minority-Group Relation*. New York: John Wiley and Sons.

Bursztyn, L., T. Chaney, et al. (2024). "The Immigrant Next Door". In: *American Economic Review* 114(2), pp. 348–384.

Dustmann, C., K. Vasiljeva, et al. (2019). "Refugee Migration and Electoral Outcomes". In: *The Review of Economic Studies* 86(5), pp. 2035–2091.

Facchini, G. and A. M. Mayda (2008). "From Individual Attitudes towards Migrants to Migration Policy Outcomes: Theory and Evidence". In: *Economic Policy* 23(56), pp. 652–713.

Facchini, G. and A. M. Mayda (2012). "Individual Attitudes Towards Skilled Migration: An Empirical Analysis Across Countries". In: *The World Economy* 35(2), pp. 183–196.

Hainmueller, J. and D. Hangartner (2013). “Who Gets a Swiss Passport? A Natural Experiment in Immigrant Discrimination”. In: *American Political Science Review* 107(1), pp. 159–187.

Hainmueller, J. and D. J. Hopkins (2014). “Public Attitudes Toward Immigration”. In: *Annual Review of Political Science* 17(1), pp. 225–249.

Hainmueller, J. and D. J. Hopkins (2015). “The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes toward Immigrants”. In: *American Journal of Political Science* 59(3), pp. 529–548.

Mayda, A. M. (2006). “Who Is Against Immigration? A Cross-Country Investigation of Individual Attitudes toward Immigrants”. In: *Review of Economics and Statistics* 88(3), pp. 510–530.

Steinmayr, A. (2021). “Contact versus Exposure: Refugee Presence and Voting for the Far Right”. In: *The Review of Economics and Statistics* 103(2), pp. 310–327.

Chapter 1

Estimating the impact of xenophobia on immigration: evidence from terrorist attacks

Abstract

This paper analyses the sensitivity of potential migrants to anti-immigration attitudes in destination countries. We use migration plan data from the Gallup World Polls and measure anti-immigration attitudes from the Eurobarometer. To overcome important endogeneity issues, we instrument attitudes towards immigration with the number of citizens of a European country killed in terrorists attacks outside Europe. We find that terrorist attacks tend to increase negative attitudes with respect to immigration in the origin country of the victims, which validates the strength of our instrument. Those attitudes then decrease the country's attractiveness for potential immigrants, no matter their skill level.

1.1 Introduction

Over the last decades, the world has experienced a rise in anti-globalization activism and an increase in populist political attitudes in many developed countries. Recent events like the Brexit in 2020 or the election of Donald Trump reflect a deep rejection from a part of the population of the process of internationalization of relations between countries. These evolutions pertain both to a strong opposition to the exchange of goods across countries – trade – and the mobility of people across borders – migration. The rise of concerns about immigration is also reflected by a change in reported attitudes of the native population with respect to the perceived impact of the immigrants and the desirability to curb the number of arrivals of individuals from above. Over time, the increase in the anti-immigration attitudes is reflected in the voting patterns of the resident population for anti-immigration parties as illustrated by the last round of European elections in June 2024.

A very large literature in social science has focused on the determinants of anti-immigration attitudes[1]. This large literature has analyzed the role of various factors, and in particular economic and cultural ones. While economic factors like those related to the perceived impact of immigration of the domestic labour market play an obvious role in shaping the attitudes (Facchini and Mayda, 2012), cultural factors are definitely important as well (Card et al., 2012). A less studied question is the impact of these attitudes on different outcomes. In this paper, we focus on an important outcome, the level of attractiveness of the country for potential migrants. In particular, we empirically address the following question: to what extent a rise in anti-immigration attitudes affect the location choices of international migrants?

One of the potential reasons for the relative lack of compelling evidence about this question is related to the issues concerning the identification of the impact of these attitudes on migration plans or mobility outcomes. One major issue is the presence of confounding factors of migration desires that are correlated with the anti-immigration attitudes in the potential destination. The sign and magnitude of such a correlation is ex-ante ambiguous. On the one hand, along the contact theory, higher presence of immigrants might decrease the racial prejudice of natives. On the other hand, it could be the case that an increase in the attractiveness of a specific destination also creates concerns on the side of its native population about immigration, which in turn boosts anti-immigration attitudes. Such a pattern is consistent with the observed rise of anti-immigration attitudes in two major hosting European countries in the aftermath of the refugee crisis in 2015. [2] In any case, the presence of confounding factors creates a potential bias in the estimation of the effect of attitudes and in turn requires

[1] See for instance Alesina and Tabellini (2024) for a survey of this literature.

[2] Between 2014 and 2015, following the Euroborameter, the increase in share of the native German and Swedish population raising concerns about immigration doubled in both countries.

the use of a sound identification strategy. This is the major contribution of this paper.

In this paper, we assess empirically the impact of attitudes on the attractiveness of the European destinations for potential migrants. Attractiveness of a location is captured by the share of respondents in the Gallup World Poll survey, a comprehensive and harmonized survey across countries, planning to migrate to that location. We measure anti-immigration attitudes from the Eurobarometer. Our estimations aim at testing whether an increase in negative attitudes about immigration in a country deters potential migrants from choosing this country as their preferred location. In order to identify this effect, we propose a new instrumental variable for attitudes in a country, namely the number of its citizens killed in a terrorist attack outside the country or the region of the country. The idea of this instrument is that, unlike attacks taking place in the country, such events induce a more or less random distribution in the number of victims by nationality and therefore do not affect the attractiveness of the country for potential migrants. In contrast, this type of event might affect the way natives perceive future immigrants and hence might change their attitudes towards immigration. We find indeed that the number of citizens killed outside Europe significantly boosts negative attitudes about immigration in the country of the victims.

Our main results are the following ones. Terrorist attacks outside the country or even outside Europe making casualties among citizens lead natives to adopt more negative attitudes with respect to immigrants. Second, using our IV approach based on this, we find that negative attitudes of native residents decrease the attractiveness of the country for potential immigrants. This result is not obtained when running naive regressions that are subject to endogeneity issues. Negative attitudes affect in a similar way the attractiveness of a country across all categories of skilled immigrants. Our results are robust to alternative definitions of the instrument and are confirmed through a placebo analysis. All in all, our results conclude in the existence of global negative impact of xenophobia on the attractiveness of a country.

Our paper contributes to three strands of the literature.

First, we contribute to the literature looking at the impact of attitudes on immigration. While there is an extensive literature on the determinants of attitudes, there are only a few papers exploring their effects on immigration outcomes. Using a dyadic framework, Gorinas and Pytliková (2017) study the impact of anti-immigration attitudes on immigration flows in 30 OECD countries over the 1980-2010 period using information from the Integrated Values Survey. They find a negative impact of specific measures such as the propensity of natives to discriminate immigrants or the reluctance to have a foreign neighbor. Nevertheless, while they control for a large set of fixed effects, they do not account explicitly for the endogeneity nature of natives' attitudes. Iasio and Wahba (2023) analyze the same effect for 21 European destinations but account for the endogeneity of attitudes using a measure of natives' cultural conformity based on religious practice of low educated natives. They also find a negative impact on immigrants inflows. At the micro-level, Slotwinski and Stutzer (2019) study the effect of the vote against minaret construction in Swiss municipalities, and find that migrants are less likely to move to these municipalities. Even if it is focused on Muslim migrants, this policy discouraged non-Muslim migrants to come, and this effect is stronger for high-skilled migrants. A related literature looks at the impact of integration policies on the attractiveness of the destinations for potential migrants. It can be expected that, over the long run, integration policies reflect natives' attitudes on immigration. In that respect, Beine et al. (2020) find a negative impact of labour market integration policies in OECD countries on the willingness of individuals to migrate to that destination. Nevertheless, their empirical analysis fail to explicitly account for the endogenous nature of these policies.

In this paper, we contribute explicitly to this literature by proposing an identification strategy allowing to capture the causal impact of natives' attitudes with respect to immigration on the attractiveness of the destination. Beyond this explicit aim, we also extend the literature by looking at the impact on stated preferences rather on the actual outcomes. Actual movements are also driven by constraints faced by the individuals, which might also be correlated with attitudes. By using migration plans, we emphasize the role of natives' attitudes as a self-selection factor of location choice.

Secondly, this paper also speaks to the large literature on the determinants of international migration. Various "push" and "pull" factors are found to affect migrants' destination choice. An obvious one is the GDP per capita at destination, which proxies the income that one could gain by migrating there (Ortega and Peri, 2013). Higher GDP per capita at destination increases migration flows to this destination. Another important factor is the size of network in the destination country. This can take the aspect of an important diaspora (Hatton and Leigh, 2011; Docquier et al., 2014; Manchin and Orazbayev, 2018) or of personal connections (Bertoli and Ruyssen, 2018).

While the "push" and "pull" factors have been extensively studied, our paper is more related to a complementary pair that has received less attention: the "retain" and "repel" factors (Arango, 2000; Schewel, 2020). "Retain" refers to the factors that make someone stay in their home country because of the benefits of immobility. In particular, having some family, having children, and being involved in the local community are important reasons of why one does not want to move (see e.g Ritchey, 1976; Mincer, 1978; Ermisch and Mulder, 2019). "Repel" refers to the factors that discourage someone from going to a particular destination. For instance, conflict in one's home country acts as a "push" factor, but conflict in another country acts as a "repel" factor, as it discourages people from migrating there. In our paper, anti-immigration attitudes play the role of repel factors as they discourage people from migrating.

Finally, we contribute to the literature on the relationship between terrorism and attitudes towards migrants. In particular, the results of our instrumental variable corroborate those of several other papers showing that terrorism leads to more anti-immigration attitudes[3]. In particular, Cruz et al. (2020) find that terrorism incidents in a country lead to an increase in anti-immigration sentiment, in particular towards out-group migrants. While we could expect the effect to be higher when terrorist attacks happen directly in the country, those can also have an international impact. Bove et al. (2021) find that attacks happening in neighboring countries lead to more restrictive migration policies due to the common perception of causality between migration and terrorism[4]. Similarly, Legewie (2013) finds that a terrorist attack highly covered in the news may affect attitudes towards migrants, even if it happens on the other side of the planet. We add to this literature by bringing a new dataset covering the number of European victims of terrorist attacks happening outside Europe, broken down by nationality. This data mostly comes from news press agencies articles and was checked against the Global Terrorism Database, which is usually used in this literature.

The rest of the paper is organized as follows. Section 1.2 describes the empirical strategy, including the endogeneity issue and the description of our instrumental variable; section 1.3 details the various data sources used; section 1.4 presents the results; and section 1.5 concludes.

1.2 Estimating the impact of attitudes on migration aspirations

1.2.1 Empirical framework

The model that we use is based on a traditional micro-founded gravity equation, which has been widely used to estimate the determinants of bilateral migration flows and aspirations (see among others Beine et al. (2016) and Head and Mayer (2014) for a detailed explanation on the use of gravity equation in the migration literature). [5]

[3] See Helbling and Meierrieks (2020) for a comprehensive review.

[4] It is important to note here that whether this causality exists or not is irrelevant to using terrorism to instrument attitudes. The strength of this instrument relies on the fact that there is a belief of this causal relationship.

[5] This model might be derived from the equilibrium of a Random Utility Model (RUM). See Beine et al. (2011) or Grogger and Hanson (2011) as examples. We skip here this derivation to focus on the econometric specification.

The structural equation of the model that we bring to the data takes the following form:

$$M_{ijt} = \exp(\gamma_{ij} + \gamma_{it} + \beta_1 Att_{j,t-1} + \mathbf{X_{jt}}'\lambda)\, \varepsilon_{ijt} \tag{1.1}$$

where M_{ijt} is the share of respondents in country i at time t that aspire to emigrate from country i to country j among all respondents surveyed at time t. We measure M_{ijt} as:

$$M_{ijt} = \frac{Plan_{ijt}}{\sum\limits_{j \neq i} Plan_{ijt}} \times \frac{Yes_{it}}{Resp_{it}} \tag{1.2}$$

where $Plan_{ijt}$ is the number of respondents having a plan to migrate from country i to country j at time t. $\frac{Yes_{it}}{Resp_{it}}$ gives the proportion of respondents in wave t that state a willingness to emigrate internationally[6].

Att_{jt-1} is a variable measuring negative attitudes with respect to immigrants at time $t-1$. Attitudes are lagged to ensure that migration plans are expressed after the observation of these attitudes by natives of country j. X_{jt} is a vector of control variables affecting the attractiveness of destination j.[7] In this specification, we control through fixed effects γ_{ij} and γ_{it} for unobserved dyadic and time-varying origin country specific factors.

Since GWP data include a large set of individual characteristics of the respondents, it is possible to focus on a sub-sample of the population at origin, for instance in terms of education level. This allows to consider heterogenous effects of anti-immigration attitudes across potential migrants. Similarly, since GWP data gathers most origin countries in each wave, it is possible to breakdown analysis by type of country of origin.

The estimation of equation (1.1) is subject to some specific issues. First, a large proportion of M_{ijt} is made of zeroes, reflecting that migration plans are concentrated on specific destination countries.[8] As identified by a couple of papers (Santos Silva and Tenreyro, 2006; Santos Silva and Tenreyro, 2011; Correia et al., 2020), such a high presence of zeroes leads to two specific issues, namely selection and bias in the estimator of the parameters of equation (1.1). To address these issues, it is advised to consider an exponential form of the equation,

[6] Therefore, our dependent variable is the respondents having emigration plan to a specific destination as a share of total respondents, including the intended stayers.

[7] $\mathbf{X_{jt}}'$ includes the GDP per capita, the population size, an index representing the possibility for migrants to enter the labor market at destination, and the number of victims of terrorism in the destination country j at year t.

[8] In our sample, we observe that about 90 percent of our observations are zeroes. This high concentration of choices on a limited set of destinations is a strong stylized fact on intention data (see Docquier et al., 2014 on this) and generally speaking on dyadic international movements of individuals.

which can be estimated by the Poisson Pseudo-Maximum Likelihood (PPML) estimator.

1.2.2 Endogeneity of attitudes

A more serious econometric problem is the endogeneity of attitudes in equation (1.1). Our variable of interest Att_{jt-1} is indeed likely to be correlated with unobserved factors that also affect migration plans. For instance, a generous social security system at destination could encourage migrants to come, but can also increase anti-immigration attitudes if natives consider that foreigners do not contribute enough to this system while they benefit from it. Another example is provided by the massive arrivals of Syrian refugees in some countries in 2015. The liberal position of Angela Merkel in Germany induced a large inflows of more than one million of Syrian refugees but at the same time also triggered some rise in the anti-immigration attitudes of a subset of the German population. This reflects a more general phenomenon in the hosting countries [9].

The endogenous nature of attitudes calls for a specific econometric treatment of equation (1.1). In this paper, we provide an instrumental variable solution to address this issue. In our context, such an instrument should predict anti-immigration attitudes of natives without impacting directly migration plans. Unfortunately, instruments based on lagged values of attitudes are likely to be invalid since unobserved factors of attractiveness of the destination are likely to be persistent over time. This calls for the use of an external instrument. The main contribution of this paper is to propose such an instrument. This is, to the best of our knowledge, the first paper to provide such a solution.

1.2.3 Victims of terrorist attacks as an instrument

We rely on an instrument based on a specific measure of terrorist attacks, namely the number of citizens of country j killed in terrorist attacks outside Europe in a given period. The idea of such an instrument is that terrorist attacks making victims from j abroad either do not target this destination or generate a more or less random distribution of victims by nationality. A good example of such an attack is provided by the bombings in night clubs in Bali, Indonesia in 2002 that led to 202 victims. While these attacks explicitly target foreigners in Indonesia, the number of victims by citizenship was likely to be unknown by the perpetrators of these attacks and has a random component.

One could argue that terrorist attacks taking place abroad could predict the number of terrorist attacks on the destinations' soil and affect the attractiveness of the destination,

[9] See FOURN Léo, "L'installation des réfugiés syriens en Europe face au recul de l'hospitalité", Migrations Société, 2018/4 (N° 174), p. 17-31. DOI : 10.3917/migra.174.0017. URL : https://www.cairn.info/revue-migrations-societe-2018-4-page-17.htm

violating the exclusion restriction. To address this important point, we do two separate things. First, we control explicitly for terrorists attacks taking place in country j. A recent literature (Foubert and Ruyssen, 2024) has looked at the potential impact of such events on migration intentions. They find that terrorist attacks exert a negative impact on immigration flows or intentions. In general, we find similar effects of such attacks on migration plans, albeit without accounting for the potential endogeneity of such events.

Second, our instruments do not include nationals killed in terrorist attacks on the national territory since this could affect the perceived attractiveness by potential migrants. In our benchmark estimations, we take in this respect a very conservative approach by excluding terrorist attacks taking place not only in neighbouring countries but in Europe as a whole. One could argue that through some contagion, terrorist attacks taking place, say, in Italy or in Norway could raise the probability of future attacks in, say, France, which would indirectly affect French attractiveness.[10] Our identification therefore assumes that while potential migrants might account for terrorist attacks at destination in choosing their optimal location, they do not account for terrorist attacks killing nationals of the country when these take place far away from this destination.

We use the number of victims in attacks outside Europe to capture the possible impact these events have on natives' opinions and attitudes regarding immigration. Our instrument turns out to be a reasonably strong predictor of the anti-immigration attitudes. These results are in line with a recent literature showing that terrorism increases anti-immigration attitudes (Legewie, 2013; Nussio et al., 2019; Ferrín et al., 2020), even when attacks happen abroad (Böhmelt et al., 2020). The channels of influence of such events on opinions are multiple but a usual suspect is the coverage by the national press. The magnitude of press coverage is likely to be proportional to the severity of the events for the citizens of the country. This severity might be better proxied by the number of casualties rather than a simple dummy capturing the mere occurrence.

Finally, we ensured that the timing of the attacks and of our measure of attitudes are consistent. Attitudes are measured from the Eurobarometer surveys (see section 1.3.2), which usually take place over one or two weeks at different times of the year depending on the country. We first compute the number of victims of terrorism in the year preceding the Eurobarometer interview. This gives us an individual measure that we then average over each country-wave.

To sum up, we find that our instrument is a good predictor of attitudes at destination and are confident its specific nature leads to comply with the exclusion restriction of the IV

[10] In robustness checks, we also show that our results hold when including attacks either in all other European countries or in non neighbouring European countries. See section 4.2.

procedure.

1.2.4 Control Function Estimation

We use the control function approach, which might be seen as the counterpart of the IV estimation for non linear models such as the exponential one (Wooldridge, 2014). [11] It involves two steps. In the first step, we use linearly regress the endogenous variable on all the controls and the relevant fixed effects. In the second step, we estimate the structural equation (1.1), adding the residuals of the first step regressions. The underlying idea is that these residuals capture the role of unobserved factors of attractiveness and corrects the estimation for the its endogeneity nature. This method can be expressed as follows:

$$Att_{jt} = \lambda_j + \lambda_t + \alpha_1 Terrorism_{jt} + \mathbf{X_{jt}}'\delta + \nu_{jt} \tag{1.3a}$$

$$M_{ijt} = \exp(\gamma_{ij} + \gamma_{it} + \beta_1 Att_{j,t-1} + \mathbf{X_{jt}}'\lambda + \beta_2 \hat{\nu_{jt-1}}) + \varepsilon_{ijt} \tag{1.3b}$$

where equation (1.3a) is the first step regression and equation (1.3b) is the structural equation to estimate. To ensure the reliability of the standard errors in the second step, we bootstrap this two-step procedure, as recommended in the literature (Wooldridge, 2015; Lin and Wooldridge, 2019). We use bootstrap on the full sample with replacement clustered by origin-destination dyad. We use 1000 repetitions for each two-step procedure.

As emphasized by Wooldridge (2014), an attractive feature of the Control Function (CF) approach is that it provides a kind of Hausman test of the endogeneity of the variable of interest. A significance of the parameter β_2 would tend to suggest that the variable is endogenous in equation (1.1). Furthermore, the sign of $\hat{\beta}_2$ is indicative of the impact of such an endogeneity on the estimated coefficient and the direction of the estimation bias relative to equation (1.1).

1.3 Data

In order to estimate equations (1.3a) and (1.3b), we need to measure three main variables: bilateral migration plans, attitudes with respect to immigration and a measure of European victims by citizenship killed in terrorist attacks by location. Our final dataset comprises 149 origin countries, 33 destination countries, and goes from 2010 to 2015. In this paper, we focus on European destinations due to data availability with respect to attitudes towards immigrants.

[11] This method has been for instance used by Miroudot and Rigo (2021) to estimate bilateral trade flows.

1.3.1 Migration plans

Our dependent variable is derived from individual migration plans. These are retrieved from the Gallup World Polls (GWP), an annual worldwide comprehensive survey that gathers data of a large number of respondents in each country. An attractive feature of the GWP is that it is harmonized across countries, which makes a cross-country investigation possible. Depending on the size of the country, the number of respondents vary from 500 individuals for small countries like Luxembourg to 1000 for most countries and 3000 for large countries (like the US or China). Our unit of analysis involves a specific country of origin, a specific destination and a specific year. Our measure of migration plan is aggregated from the individual data. We make use of the reported education level of the respondent to capture possible heterogeneous effects of attitudes on migration plans.

In order to measure migration plans, we use two questions that involve emigration plans and optimal destination choices. First, respondents are asked "Ideally, if you had the opportunity, would you like to move permanently to another country, or would you prefer to continue living in this country?". The second daughter question is asked only to those having answered positively to the first one and is stated as: "To which country would you like to move?". These two questions lead to what is called migration aspirations, which are extensively used in the migration literature. These aspirations reflect mobility preferences that are unconstrained. They are not backed up by any deed or project, which has induced scholars to use alternative measures of intentions (see for instance Clemens and Mendola (2020)).

In this paper, we use a follow-up question about more tangible plans: "Are you planning to move permanently to another country in the next 12 months, or not?".[12] While on average 22.3 % of the respondents would like to move permanently in an ideal world, only about between 2 and 3% of the respondents worldwide have plans to do so in the following year. Of course, the magnitude depends on the area of origin, with African respondents expressing the highest level of desired emigration. Plans are used as measures of intentions rather than aspirations. We aggregate the answers at the national level to compute bilateral intended migration flows. Figure 1.B.1 in the online appendix shows the evolution in the share of respondents with an intention to emigrate, i.e. the proportion of those having plans to emigrate in the next 12 months.

One might be concerned that plans of migration may not materialise and therefore would not reflect actual migration flows. However, attractive locations reflected by the mobility preferences in the GWP are those that receive the highest number of international immigrants. It is important to note that about 90% of bilateral flows are zeroes. This is a common feature in

[12] This question was only present in the Gallup interviews conducted between 2010 and 2015. Table 1.A.2 in the Appendix shows the coverage per country and per wave.

bilateral data of trade and migration. In the case of bilateral migration plans, the proportion of zeroes is even higher since preferences are slightly more concentrated on a few destinations compared to actual movements. Table 1.A.1 in the online appendix shows the preferred destinations on average for each year over our period of study (with Home being the favourite destination for those who don't have any migration plans in the following year).

1.3.2 Attitudes towards immigration

Our interest variable is the attitudes towards immigration in destination countries. In this paper, we focus on European countries and we use the Eurobarometer, a survey conducted on a yearly basis.[13] We focus on the native respondents only, i.e individuals born in the country where they live. We use the question "What do you think are the two most important issues facing [our country] at the moment?". Respondents are given a list of 14 issues including immigration. While answering "immigration" does not stricto sensu mean that one is against immigration, we consider that people with anti-immigration attitudes are more likely to mention it. This measure is already used in the literature to capture anti-immigration attitudes. (see e.g Böhmelt et al., 2020; Hatton, 2021). We use the share of individuals that chose immigration as our measure of anti-immigration attitudes.

In the Eurobaromater, attitudes are usuall collected in May-June, with the precise dates varying from country to country. In contrast, the collection by Gallup of migration aspirations varies across countries. GWP surveys are conducted all year long. Therefore, we use lagged values of attitudes in equation (1.1) to avoid using migration plans measured before attitudes towards immigration.

1.3.3 Terrorism

In equation (1.3a), our instrument is the number of citizens of each destination country that were killed each year in terrorist attacks outside of Europe. Other studies using terrorism data usually rely on the Global Terrorism Database (GTD, Enders et al., 2011), which collects information on more than 200,000 terrorist attacks from 1970 onwards. However, this database does not provide the nationality of the victims of terrorism, except for US citizens. Therefore, to obtain this piece of information, we used the online service Factiva to search for news coming from the main news agencies (Reuters, Associated Press, Agence France Presse). Specifically, we searched for several keywords associations and manually reviewed the articles returned by Factiva to extract the number and nationality of victims of terrorism in the world. With

[13] An alternative would be to use questions drawn from the European Social Surveys (ESS). Nevertheless, the bi-annual frequency as well as a significant share of missing data for specific countries would induce a much small and selective sample of destination countries over our investigation period.

this information, we could then get the number of citizens of European countries that were killed in terrorist events outside of their country each year. Because this information is very specific and articles all have different formulations, we couldn't rely on an automated process. For English-speaking news agencies (Reuters and AP), we searched for articles containing the word "citizen" and at least one of the words "killed", "assassinated", "beheaded", and "shot". For Agence France Presse, we searched for the keywords "ressortissant tué" or "citoyen tué" (*killed national* or *killed citizen*).

One may be concerned that these three news agencies essentially focus on the United States and Western Europe, and are less likely to report attacks on citizens from Eastern European countries. To lower the risk of under-representation of Eastern European countries, we also searched for articles coming from the Baltic News Network and Sofia Press Agency, two press agencies with international coverage implemented respectively in Austria and in Bulgaria. [14] Although these news agencies provide less content, they reduce the risk of missing attacks concerning Eastern European countries. In total, we reviewed more than 5,000 articles over the period 2006-2019 to gather data on the number and nationalities of victims of terrorism abroad.

In most cases, several articles from different news agencies reported the same information, which lowers the risk that some events were forgotten. News agencies also quickly update their reports so that the number of victims is accurate. When possible, we also checked the information from Wikipedia pages about terrorist attacks, which were written some time after and therefore have more complete information. We only kept attacks that were also referenced in the Global Terrorism Database to ensure that the events we classified are terrorist events, and not unrelated accidents.

Finally, since we consider terrorist attacks as predictors of attitudes, we have to ensure that those attacks happened before the surveys took place. The Eurobarometer is usually conducted over several weeks of May-June. Therefore, not all respondents were exposed to the same attacks. To address this, we first computed the number of attacks that happened in the past year relative to the interview date, hence obtaining a measure of exposure to terrorism at the individual level. We then averaged this result at the country-year level. Figure 1.1 shows the aggregate yearly number of victims between 2007 and 2018. It displays an important variation between years, which is important to ensure the strength of our instrument.

[14] Note that the Baltic News Network was created in 2010.

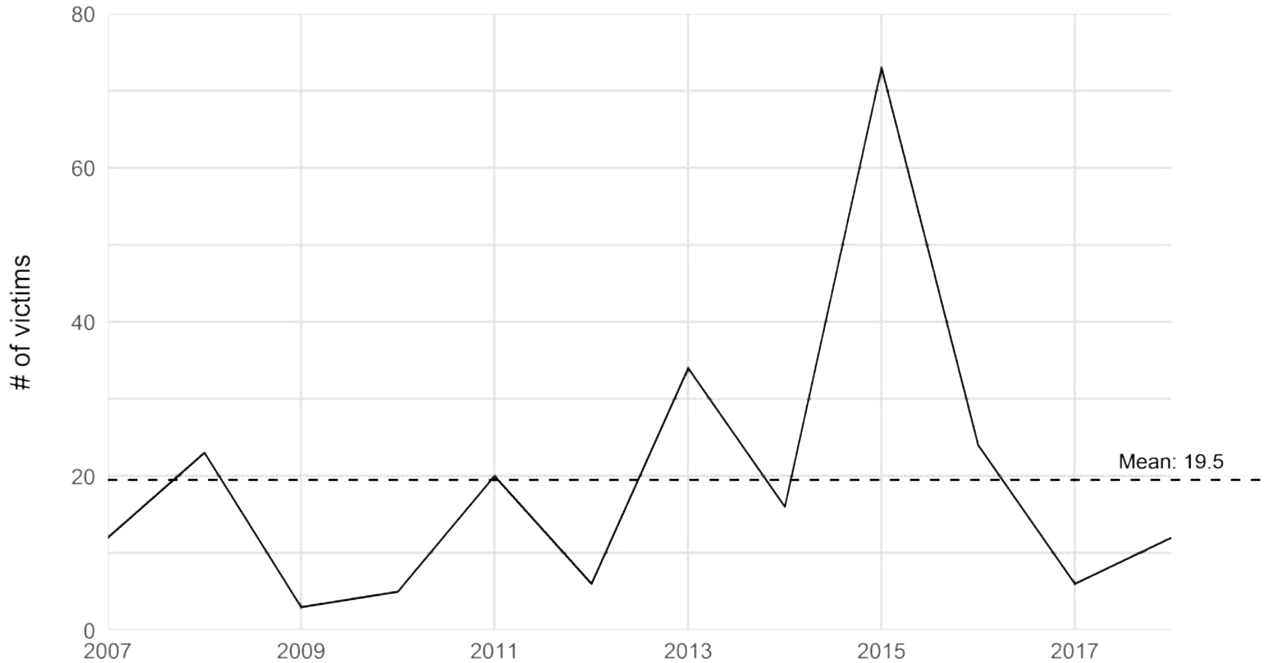


Figure 1.1: Total number of European victims of terrorist attacks outside Europe

1.3.4 Other data

We use several control variables in our main specification. We first capture variation that are destination specific and changing over time. We include the GDP per capita in the destination country, as it is an important pull factor. We also proxy the size of the destination country by its population size. GDP per capita and population size data come from the World Development Indicators (World Bank, n.d.). We also capture the role of policies at destination.

An important control is the number of terrorism victims of attacks taking place on the territory of the destination country. As explained before, this has been found in previous work to be a (negative) factor of attractiveness for potential migrants. It is also an important control shutting down one specific channel through which our instrument could be related directly to migration plans. This control is built using the Global Terrorism Database (GTD).

Finally, we saturate our specification with origin-time fixed effects to control for push factors in origin countries, such as political instability, conflicts or adverse climatic conditions, as well as origin-destination fixed effects to control for time-invariant characteristics of country pairs, such as the distance or existence of past colonial relationship.

Table 1.1: Impact of anti-immigration attitudes on migration plans: PPML

	(1) Mig	(2) Mig (HSMS)	(3) Mig (HS)	(4) Mig (MS)	(5) Mig (LS)
Eurobaro., lag	0.019***	0.017***	-0.001	0.026***	0.009
	(0.006)	(0.006)	(0.015)	(0.007)	(0.013)
GDP pc (log, lag)	2.018*	2.800**	6.010**	2.801	3.860
	(1.214)	(1.422)	(2.835)	(1.767)	(2.519)
Unemp. rate (lag)	-0.009	-0.010	-0.087	0.005	0.043
	(0.025)	(0.031)	(0.063)	(0.038)	(0.057)
Terrorim at dest., lag	0.009***	0.004	-0.002	0.003	0.020***
	(0.003)	(0.004)	(0.008)	(0.005)	(0.005)
Pseudo-R2	0.376	0.408	0.419	0.400	0.324
Origin-Year FE	Yes	Yes	Yes	Yes	Yes
Origin-Destination FE	Yes	Yes	Yes	Yes	Yes
Observations	4651	4307	1851	3474	1234

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Heteroskedasticity-robust standard-errors in parenthesis. Variable Mig. is the ratio of movers from i to j over stayers derived from the standard RUM model. LS, MS, and HS correspond respectively to low-skilled (up to 8 years of education), middle-skilled (9-15 years of education), and high-skilled (4 years of education beyond high school). HSMS includes HS and MS individuals. Variable Eurobaro. is the share of individuals in the destination country that think immigration is one of the main two issues in their country. Control variables are the log of GDP per capita, the unemployment rate, and the number of victims of terrorist attacks in the destination country lagged by one year.

1.4 Results

1.4.1 Baseline results

Table 1.1 reports the PPML estimations of equation (1.1) without using our instrumental variable. We consider all types of respondents (column 1) as well as potential migrants for different level of education (columns 2-5). The results confirm in an unambiguous way that estimation of equation (1.1) is subject to significant endogeneity issues. Anti-immigration attitudes are found to have either no effect on the attractiveness of the destination, or even a positive one. While the former result could be rationalized theoretically, the latter one is fully counter-intuitive. This suggests the existence of a positive selection bias in the estimation. One example of a specific pattern generating this bias is that a higher (unobserved) level of attractiveness in a country leads its natives to be relatively more against immigration. Such a pattern is for instance consistent with the perception of a welfare magnet of immigration by natives.

Table 1.2 reports the results of the control function estimations dealing with the presence of this endogeneity issue. In Panel A, we report the first-stage results, i.e. the impact of victims of terrorist attacks outside Europe on anti-immigration attitudes. For all specifications, we show that attitudes of respondents towards immigrants become more negative when countrymen are killed in terrorist attacks. F-stat values suggest that the instrument is a strong predictor of these negative attitudes.

Panel B reports the control-function estimation of equation (1.3b). In contrast with the OLS results of Table 1.1, we find a negative and significant effect of attitudes on the propensity of choosing the destination by potential immigrants. In all specifications, this impact is found to be negative. Furthermore, the estimate of the first-stage residuals is very significantly positive, confirming that naive regressions of equation (1.1) are subject to a significant bias. The inclusion of the first-stage residual captures the impact of unobserved factors of attractiveness (such as the welfare magnet or quality institutions). Its inclusion in equation (1.1) breaks down the correlation of these factors with the error term that generated the bias in estimations of Table 1.4.1. The significantly positive coefficients of the first-stage residuals suggests a positive omitted variable bias in naive regressions. Such a bias is consistent with the idea that natives from relatively attractive countries tend to form relatively more negative attitudes with respect to immigrants, for instance for fear that the latter will benefit from a generous redistribution system.

In terms of control variables, the results in Table 1.2 suggest a positive impact of income at destination and a negative impact of unemployment rate at destination. We do not find a negative impact of terrorist attacks on attractiveness of the country, in contrast with Foubert and Ruyssen (2022).

1.4.2 Heterogeneity analysis

Heterogeneity by education levels

An important source of heterogeneity of the impact of attitudes on attractiveness is the level of education of potential migrants. In the extensive literature on the determinants of human mobility, education level is often a major source of differences in sensitivities to the traditional factors. This pertains to the impact of distance, networks, revenues at destinations and origin, to quote just the most important ones. Regarding how attitudes are internalised by potential migrants, such an heterogeneity could be driven by the purpose of the migration project. It can be argued for instance that unskilled migrants come mainly for economic reasons and pay less attention to attitudes of natives. In the same vein, it could be argued that high skilled migrants pay more attention to attitudes since they care more about their assimilation. On the other hand, given the type of occupations and given they are more substitutes to native workers, unskilled immigrants are more likely to be subject to discrimination, due to their specific professional occupation and their higher capacity to interact with natives.

A higher sensitivity of highly skilled immigrants could give rise to a vicious circle. A higher proportion of low-skilled immigration could indeed fuel negative attitudes of natives, which is turn would deter more high-skilled immigrants to come, lowering the skill content of future flows.[15] In order to assess such a case, we carry out estimations by education level.We use 3 standard education levels: primary (defined as low-skilled) , secondary and non college higher education level (middle-skilled), and college-educated (high-skilled). Table 1.B.1 in the Appendix reports the results. Two comments are in order. First, we find compelling evidence that negative attitudes affect the willingness to come for all types of migrants, including low-educated ones. This result is for instance in line with Beine et al. (2024) who found that a Trump reelection in 2025 would decrease the perceived attractiveness of the US for all types of migrants. Second, we do not find evidence in favour of a vicious circle, i.e. higher sensitivity for high-skilled immigrants compared with low-skilled one. Estimates of attitudes are not different across education levels at conventional significant levels. It should be nevertheless stressed that this last result is obtained from a sample with a lower number of observations.[16]

Heterogeneity by country of origin

We also consider some heterogeneity by type of origin of the respondents. We first break down our investigation by level of development in the origin country. We use the World Bank classification of countries in terms of development. High income countries of origin refer here to the so-called North-North migration pattern while migrants coming from low and middle-income countries are involved more in the so-called South-North migration corridors. Table 1.B.2 in the Appendix reports the results, for all types of migrants as well as by education level. We find no evidence of heterogeneity in the effects of attitudes.

Second, we look at whether European aspirational immigrants are more or less sensitive to anti-immigration attitudes. It could be argued that due to the higher proximity, respondents are better informed about these attitudes and take this more into account. Similar to the heterogeneity analysis by level of development, we do not find any evidence of a specific effect for respondents coming from another European country.

[15] References to some existing evidence on vicious circle

[16] This happens when we focus on respondents of a particular skill level because the number of respondents who have migration plans decreases and less non-zero origin-destination corridors remain. Therefore, cases where origin-destination corridors only have 0 flows on the full period are more frequent, and those are completely absorbed by origin-destination fixed effects and are dropped from the estimation.

Robustness check: alternative measures of instrument

The building of our instrument used so far aims at complying with the exclusion restriction of the control function procedure. The exclusion of victims in terrorist attacks not only on the territory of the destination but also in all other European countries rules out any confounding effect that could come from contagion effects in the perception of potential immigrants. It could be nevertheless considered that these contagion effects are very weak, if not non existing in the way terrorist attacks are perceived by respondents.

Therefore, in Table 1.3, we assess whether our results are robust to some alternative definitions. In column (1), we include, on top of victims killed outside Europe, nationals deceased in terrorist attacks located outside the territory and its neighbouring countries. In column (2), we include also in the instrument nationals killed in neighbouring European countries. Estimates from table 1.3 show that both first-stage and second-stage results are robust to these alternative definitions. Both alternative instruments are strong predictors of anti-immigration attitudes and estimates from the structural equation of attitudes are significantly negative, in line with findings from table 1.2.

Placebo: using future attitudes

Finally, we conduct a placebo analysis to assess further the robustness of our results. We look at the impact of future attitudes on migration plans. More specifically, we use attitudes collected in the Eurobarometer two years after the elicitation of location choices in the GWP.[17]

In table 1.4, we conduct CF estimations, for all migrants altogether and distinguished by education levels. Results show that, unlike past anti-immigration attitudes, future attitudes by native have no impact on the perceived attractiveness of potential immigrants.

1.5 Conclusion

Immigration is frequently a core topic in recent political debates. The effect of immigration on natives' attitudes towards migrants has been widely studied in the two decades, and the economic and cultural determinants are both important. However, the effect of this hostility towards migrants on the decision of future migrants to go a country has received much less attention, mostly due to the endogenous relationship between attitudes towards migrants and migration flows.

[17] We use attitudes two years after rather than one year for two reasons. First, attitudes tend to be quite correlated over time. Second, because location intentions from the GWPS are measured at different moments across origin countries, using attitudes one year after would imply that, for some countries, attitudes in $t+1$ would be measured not long after the location choices.

In this paper, we used migration plan data using the Gallup World Polls surveys and measure anti-immigration attitudes from the Eurobarometer. To overcome the endogeneity issues, we instrumented attitudes in a country by the number of citizens of this country who were killed in terrorists attacks outside Europe.

Our main results are two-fold. First, terrorist attacks tend to significantly increase negative attitudes with respect to immigration in the origin country of the victims. Secondly, a higher hostility towards migrants decreases the attractiveness of the country for potential immigrants, no matter their skill level.

Table 1.2: Impact of anti-immigration attitudes on migration plans: Control-function

	Panel A: first stage			
	(1)	(2)	(3)	(4)
	Eurobaro.	Eurobaro.	Eurobaro.	Eurobaro.
# terr. victims out Eur.	0.258***	0.239***	0.179***	0.189***
	(0.016)	(0.016)	(0.016)	(0.016)
GDP per cap. (log)		16.690***	-6.630***	-4.443***
		(0.582)	(0.984)	(1.065)
Unemp. rate			-0.602***	-0.562***
			(0.018)	(0.020)
Terrorim at dest.				-0.015***
				(0.001)
Year FE	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes
Observations	31,752	31,752	31,752	31,752
KP F-statistic	256	234	124	141
	Panel B: second stage			
	(1)	(2)	(3)	(4)
	Mig	Mig	Mig	Mig
Eurobaro., lag	-0.042***	-0.068***	-0.101***	-0.099***
	(0.009)	(0.009)	(0.010)	(0.010)
1st stage resid. (lag)	0.071***	0.088***	0.120***	0.118***
	(0.005)	(0.005)	(0.005)	(0.005)
GDP pc (log, lag)		4.320***	1.831	1.219
		(1.036)	(1.737)	(1.830)
Unemp. rate (lag)			-0.078**	-0.085**
			(0.036)	(0.037)
Terrorim at dest., lag				0.008
				(0.005)
Pseudo-R2	0.376	0.376	0.376	0.376
Origin-Year FE	Yes	Yes	Yes	Yes
Origin-Destination FE	Yes	Yes	Yes	Yes
Observations	4651	4651	4651	4651

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Heteroskedasticity-robust standard-errors in parenthesis. Variable Mig. is the ratio of movers from i to j over stayers derived from the standard RUM model. Variable Eurobaro. is the share of individuals in the destination country that think immigration is one of the main two issues in their country. *# terr. victims out Eur.* is the number of citizens of the destination country killed in terrorist attacks outside of Europe in the year before the Eurobarometer interview. Control variables are the first stage residuals, log of GDP per capita, the unemployment rate, and the number of victims of terrorist attacks in the destination country lagged by one year. Non-parametric bootstrap was applied on both steps of the control function, using the full sample clustered by origin-destination, with replacement (1000 replications).

Table 1.3: Impact of attitudes: alternative instruments

Panel A: first stage	(1)	(2)
	Eurobaro.	Eurobaro.
# terr. victims out adj.	0.120***	
	(0.015)	
# terr. victims abroad		0.120***
		(0.015)
GDP per cap. (log)	-4.706***	-4.661***
	(1.063)	(1.064)
Unemp. rate	-0.569***	-0.569***
	(0.020)	(0.020)
Terrorim at dest.	-0.015***	-0.015***
	(0.001)	(0.001)
Year FE	Yes	Yes
Destination FE	Yes	Yes
Observations	31,752	31,752
KP F-statistic	62	61

Panel B: second stage	(1)	(2)
	Mig	Mig
Eurobaro., lag	-0.170***	-0.167***
	(0.010)	(0.010)
1st stage resid. (lag)	0.189***	0.186***
	(0.005)	(0.005)
GDP pc (log, lag)	0.870	0.902
	(1.829)	(1.829)
Unemp. rate (lag)	-0.126***	-0.124***
	(0.037)	(0.037)
Terrorim at dest., lag	0.007	0.007
	(0.005)	(0.005)
Pseudo-R2	0.376	0.376
Origin-Year FE	Yes	Yes
Origin-Destination FE	Yes	Yes
Observations	4651	4651

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Heteroskedasticity-robust standard-errors in parenthesis. Variable Mig. is the ratio of movers from i to j over stayers derived from the standard RUM model. Variable Eurobaro. is the share of individuals in the destination country that think immigration is one of the main two issues in their country. # *terr. victims out adj.* and # *terr. victims abroad* are respectively the number of citizens of the destination country killed in terrorist attacks outside of the country and its adjacent countries, and outside of the country only in the year before the Eurobarometer interview. Control variables are the first stage residuals, log of GDP per capita, the unemployment rate, and the number of victims of terrorist attacks in the destination country lagged by one year. Non-parametric bootstrap was applied on both steps of the control function, using the full sample clustered by origin-destination, with replacement (1000 replications).

Table 1.4: Placebo: impact of future attitudes on migration plans

	(1) Mig	(2) Mig (HSMS)	(3) Mig (HS)	(4) Mig (MS)	(5) Mig (LS)
Eurobaro., lead	-0.004	-0.012	0.028	-0.012	-0.001
	(0.009)	(0.011)	(0.028)	(0.014)	(0.019)
1st stage resid. (lead)	0.011	0.025*	-0.002	0.023	0.010
	(0.011)	(0.013)	(0.035)	(0.016)	(0.023)
GDP pc (log, lag)	1.793	2.413	4.706	2.429	3.931
	(1.849)	(2.027)	(5.699)	(2.662)	(4.652)
Unemp. rate (lag)	-0.032	-0.028	-0.102	-0.028	0.041
	(0.035)	(0.042)	(0.118)	(0.053)	(0.089)
Terrorim at dest., lag	0.009*	0.004	0.005	0.003	0.021
	(0.005)	(0.007)	(0.039)	(0.008)	(0.018)
Pseudo-R2	0.376	0.409	0.423	0.400	0.325
Origin-Year FE	Yes	Yes	Yes	Yes	Yes
Origin-Destination FE	Yes	Yes	Yes	Yes	Yes
Observations	4651	4307	1851	3474	1234

${}^{*}p < 0.10$, ${}^{**}p < 0.05$, ${}^{***}p < 0.01$. Heteroskedasticity-robust standard-errors in parenthesis. Variable Mig. is the ratio of movers from i to j over stayers derived from the standard RUM model. LS, MS, and HS correspond respectively to low-skilled (up to 8 years of education), middle-skilled (9-15 years of education), and high-skilled (4 years of education beyond high school). HSMS contains HS and MS individuals. Variable Eurobaro. is the share of individuals in the destination country that think immigration is one of the main two issues in their country. The interest variable is measured using the Eurobarometer surveys conducted two years after the Gallup surveys (which are used to compute the dependent variable). Non-parametric bootstrap was applied on both steps of the control function, using the full sample clustered by origin-destination, with replacement (1000 replications).

Appendix

1.A Additional descriptive statistics

Table 1.A.1: Attractiveness of destinations. Share of respondents (%) that have plans to migrate to [country]. If they don't have plans, then we consider 'Home' as their destination.

	2010	2011	2012	2013	2014	2015
Home	93.886	93.615	94.916	93.264	93.228	92.635
Albania	0.004	0.004	0.007	0.015	0.004	0.013
Austria	0.087	0.122	0.146	0.257	0.180	0.266
Belgium	0.101	0.121	0.105	0.149	0.164	0.187
Bulgaria	0.007	0.009	0.029	0.010	0.008	0.018
Cyprus	0.032	0.039	0.016	0.018	0.012	0.023
Czechia	0.029	0.031	0.034	0.041	0.025	0.032
Germany	0.903	0.882	0.948	1.416	1.562	1.756
Denmark	0.064	0.061	0.040	0.081	0.065	0.137
Spain	1.185	0.944	0.659	0.893	0.765	0.777
Estonia	0.002	0.001	0.002	0.006	0.001	0.005
Finland	0.024	0.070	0.065	0.069	0.080	0.082
France	1.235	1.518	0.979	0.999	1.121	1.168
United Kingdom	1.095	1.157	0.877	0.984	1.164	1.074
Greece	0.097	0.062	0.047	0.091	0.051	0.084
Croatia	0.010	0.008	0.011	0.018	0.017	0.015
Hungary	0.005	0.017	0.002	0.013	0.011	0.007
Ireland	0.060	0.071	0.036	0.048	0.050	0.062
Iceland	0.008	0.005	0.003	0.016	0.006	0.014
Italy	0.593	0.607	0.495	0.658	0.626	0.592

Lithuania	0.006	0.000	0.003	0.003	0.001	0.006
Luxembourg	0.037	0.019	0.006	0.021	0.012	0.024
Latvia	0.006	0.000	0.003	0.002	0.001	0.004
North Macedonia	0.003	0.003	0.000	0.004	0.000	0.000
Montenegro	0.002	0.003	0.001	0.000	0.000	0.003
Netherlands	0.167	0.146	0.159	0.177	0.179	0.220
Poland	0.016	0.024	0.011	0.040	0.030	0.046
Portugal	0.043	0.042	0.032	0.058	0.045	0.066
Romania	0.009	0.015	0.008	0.020	0.011	0.018
Serbia	0.002	0.001	0.029	0.031	0.037	0.017
Slovakia	0.001	0.002	0.000	0.003	0.004	0.006
Slovenia	0.007	0.003	0.018	0.009	0.011	0.016
Sweden	0.202	0.254	0.202	0.342	0.288	0.367
Turkey	0.070	0.146	0.110	0.242	0.242	0.261
Total	100	100	100	100	100	100

Table 1.A.2: Gallup coverage. Country-year that received the question 'Are you planning to move in [country] in the next 12 months?'

	2010	2011	2012	2013	2014	2015
AFG	✓	✓	✓	✓	✓	✓
AGO	.	✓	✓	✓	✓	.
ALB	.	.	✓	✓	✓	✓
ARE	✓	✓	✓	✓	✓	✓
ARG	✓	✓	✓	✓	✓	✓
ARM	✓	✓	✓	✓	✓	✓
AUS	✓	✓	.	✓	.	✓
AUT	✓	✓	.	✓	.	✓
AZE	✓	✓	✓	✓	✓	✓
BDI	.	✓	.	.	✓	.
BEL	✓	✓	.	✓	.	✓
BEN	.	✓	✓	✓	✓	✓
BFA	✓	✓	.	✓	✓	✓
BGD	✓	✓	✓	✓	✓	✓
BGR	✓	✓	✓	✓	✓	✓
BHR	✓	✓	✓	✓	.	✓
BIH	.	.	✓	✓	✓	✓
BLR	✓	✓	✓	✓	✓	✓
BLZ	.	.	.	.	✓	.
BOL	✓	✓	✓	✓	✓	✓
BRA	✓	✓	✓	✓	✓	✓
BTN	.	.	.	✓	✓	✓
BWA	✓	✓	✓	✓	✓	✓
CAF	✓	✓	.	.	.	.
CAN	✓	✓	✓	✓	.	.
CHE	.	.	.	.	.	✓
CHL	✓	✓	✓	✓	✓	✓
CHN	✓	✓	.	✓	✓	✓
CIV	.	.	.	✓	✓	✓
CMR	✓	✓	✓	✓	✓	✓
COD	.	✓	✓	✓	✓	✓
COG	.	✓	✓	✓	✓	✓
COL	✓	✓	✓	✓	✓	✓
COM	✓	✓	✓	.	.	.

CRI	✓	✓	✓	✓	✓	✓
CYP	✓	✓	.	✓	.	✓
CZE	✓	✓	✓	✓	✓	✓
DEU	✓	✓	.	✓	.	✓
DJI	✓	✓	.	.	.	.
DNK	✓	✓	.	✓	.	✓
DOM	✓	✓	✓	✓	✓	✓
DZA	✓	✓	✓	.	✓	.
ECU	✓	✓	✓	✓	✓	✓
EGY	✓	✓	✓	✓	✓	✓
ESP	✓	✓	✓	✓	.	✓
EST	.	✓	✓	✓	✓	✓
ETH	.	.	.	✓	✓	✓
FIN	✓	✓	.	✓	.	✓
FRA	✓	✓	✓	✓	.	✓
GAB	.	✓	✓	✓	✓	✓
GBR	✓	✓	.	✓	.	✓
GEO	✓	✓	✓	✓	✓	✓
GHA	✓	✓	✓	✓	✓	✓
GIN	.	✓	✓	✓	✓	✓
GRC	✓	✓	✓	✓	✓	✓
GTM	✓	✓	✓	✓	✓	✓
HKG	✓	✓	.	.	.	.
HND	✓	✓	✓	✓	✓	✓
HRV	.	.	✓	✓	✓	✓
HTI	✓	✓	✓	✓	✓	✓
HUN	✓	✓	✓	✓	✓	✓
IDN	✓	✓	✓	✓	✓	✓
IND	✓	✓	✓	✓	✓	✓
IRL	✓	✓	.	✓	.	✓
IRN	.	✓	✓	✓	✓	✓
IRQ	✓	✓	✓	✓	✓	✓
ISL	.	.	.	✓	.	✓
ISR	✓	✓	✓	✓	✓	✓
ITA	✓	✓	✓	✓	.	✓
JAM	.	✓	.	✓	✓	.
JOR	✓	✓	✓	✓	✓	✓
JPN	✓	✓	✓	✓	.	✓

KAZ	✓	✓	✓	✓	✓	✓
KEN	✓	✓	✓	✓	✓	✓
KGZ	✓	✓	✓	✓	✓	✓
KHM	✓	✓	✓	✓	✓	✓
KOR	✓	✓	✓	✓	.	✓
KWT	✓	✓	✓	✓	.	✓
LAO	.	✓	.	.	.	.
LBN	✓	✓	✓	✓	✓	✓
LBR	✓	.	.	✓	✓	✓
LBY	.	.	✓	.	.	✓
LKA	✓	✓	✓	✓	✓	✓
LSO	.	✓	.	.	.	.
LTU	✓	✓	✓	✓	✓	✓
LUX	✓	✓	.	✓	.	✓
LVA	.	✓	✓	✓	✓	✓
MAR	✓	✓	✓	✓	✓	✓
MDA	✓	✓	✓	✓	✓	✓
MDG	.	✓	✓	✓	✓	✓
MEX	✓	✓	✓	✓	✓	✓
MKD	.	.	✓	✓	✓	✓
MLI	✓	✓	✓	✓	✓	✓
MLT	✓	✓	.	✓	.	✓
MMR	.	.	✓	✓	✓	✓
MNE	.	.	✓	✓	✓	✓
MNG	✓	✓	✓	✓	✓	✓
MOZ	.	✓	.	.	.	✓
MRT	✓	✓	✓	✓	✓	✓
MUS	.	✓	.	.	✓	.
MWI	.	✓	✓	✓	✓	✓
MYS	✓	✓	✓	✓	.	✓
NAM	.	.	.	.	✓	.
NER	✓	✓	✓	✓	✓	✓
NGA	✓	✓	✓	✓	✓	✓
NIC	✓	✓	✓	✓	✓	✓
NLD	✓	✓	.	✓	.	✓
NOR	.	.	.	.	.	✓
NPL	✓	✓	✓	✓	✓	✓
NZL	✓	✓	.	✓	.	✓

PAK	✓	✓	✓	✓	✓	✓
PAN	✓	✓	✓	✓	✓	✓
PER	✓	✓	✓	✓	✓	✓
PHL	✓	✓	✓	✓	✓	✓
POL	✓	✓	✓	✓	✓	✓
PRI	.	.	.	.	✓	.
PRT	✓	✓	.	✓	.	✓
PRY	✓	✓	✓	✓	✓	✓
PSE	✓	✓	✓	✓	✓	✓
QAT	✓	.	✓	.	.	.
ROU	✓	✓	✓	✓	✓	✓
RUS	✓	✓	✓	✓	✓	✓
RWA	.	.	✓	✓	✓	✓
SAU	✓	✓	✓	✓	✓	✓
SDN	✓	✓	✓	.	✓	.
SEN	✓	✓	.	✓	✓	✓
SGP	✓	✓	✓	.	✓	✓
SLE	✓	✓	.	✓	✓	✓
SLV	✓	✓	✓	✓	✓	✓
SOM	.	.	.	.	✓	✓
SRB	.	.	✓	✓	✓	✓
SSD	.	.	.	.	✓	✓
SUR	.	.	✓	.	.	.
SVK	✓	✓	✓	✓	✓	✓
SVN	✓	✓	.	✓	.	✓
SWE	✓	✓	.	✓	.	✓
SWZ	.	✓	.	.	.	.
SYR	✓	✓	✓	✓	.	✓
TCD	✓	✓	✓	✓	✓	✓
TGO	.	✓	.	.	✓	✓
THA	✓	✓	✓	✓	✓	✓
TJK	✓	✓	✓	✓	✓	✓
TKM	.	✓	✓	✓	✓	✓
TTO	.	✓	.	✓	.	.
TUN	✓	✓	✓	✓	✓	✓
TUR	✓	✓	✓	✓	✓	✓
TWN	✓	✓	.	✓	.	✓
TZA	✓	✓	✓	✓	✓	✓

UGA	✓	✓	✓	✓	✓	✓
UKR	✓	✓	✓	✓	✓	✓
URY	✓	✓	✓	✓	✓	✓
USA	✓	✓	✓	✓	.	.
UZB	✓	✓	✓	✓	✓	✓
VEN	✓	✓	✓	✓	✓	✓
VNM	✓	✓	✓	✓	✓	✓
YEM	✓	✓	✓	✓	✓	✓
ZAF	✓	✓	✓	✓	✓	✓
ZMB	.	✓	✓	✓	✓	✓
ZWE	✓	✓	✓	✓	✓	✓

1.B Additional results for online appendix only

Table 1.B.1: Impact of anti-immigration attitudes : heterogeneity by education level

	(1) Mig (HSMS)	(2) Mig (HS)	(3) Mig (MS)	(4) Mig (LS)
Eurobaro., lag	-0.120***	-0.155***	-0.146***	-0.190***
	(0.011)	(0.030)	(0.013)	(0.026)
1st stage resid. (lag)	0.138***	0.157***	0.172***	0.199***
	(0.005)	(0.011)	(0.006)	(0.011)
GDP pc (log, lag)	1.805	4.669	1.615	2.823
	(2.002)	(5.877)	(2.546)	(4.591)
Unemp. rate (lag)	-0.101**	-0.193	-0.108**	-0.078
	(0.043)	(0.118)	(0.053)	(0.098)
Terrorim at dest., lag	0.003	-0.003	0.001	0.017
	(0.007)	(0.044)	(0.008)	(0.018)
Pseudo-R2	0.409	0.420	0.401	0.325
Origin-Year FE	Yes	Yes	Yes	Yes
Origin-Destination FE	Yes	Yes	Yes	Yes
Observations	4307	1851	3474	1234

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Heteroskedasticity-robust standard-errors in parenthesis. Variable Mig. is the ratio of movers from i to j over stayers derived from the standard RUM model. LS, MS, and HS correspond respectively to low-skilled (up to 8 years of education), middle-skilled (9-15 years of education), and high-skilled (4 years of education beyond high school). HSMS contains HS and MS individuals. Variable Eurobaro. is the share of individuals in the destination country that think immigration is one of the main two issues in their country. The instrument used in the first stage is the number of citizens of the destination country killed in terrorist attacks outside of Europe in the year before the Eurobarometer interview. Control variables are the first stage residuals, log of GDP per capita, the unemployment rate, and the number of victims of terrorist attacks in the destination country lagged by one year. Non-parametric bootstrap was applied on both steps of the control function, using the full sample clustered by origin-destination, with replacement (1000 replications).

Table 1.B.2: Result with interaction on development level of origin

	(1) Mig	(2) Mig (HSMS)	(3) Mig (HS)	(4) Mig (MS)	(5) Mig (LS)
Attitudes., lag	-0.117***	-0.125***	-0.146*	-0.149***	-0.184***
	(0.017)	(0.022)	(0.083)	(0.028)	(0.038)
Attitudes., lag × Middle Inc	0.037**	0.023	0.064	0.013	0.055
	(0.019)	(0.023)	(0.084)	(0.029)	(0.041)
Attitudes., lag × High Inc	0.004	0.008	0.049	0.005	-0.063
	(0.021)	(0.026)	(0.084)	(0.033)	(0.985)
1st stage resid. (lag)	0.113***	0.129***	0.103***	0.167***	0.165***
	(0.005)	(0.005)	(0.011)	(0.006)	(0.010)
GDP pc (log, lag)	1.485	2.073	5.742	1.749	3.306
	(1.797)	(2.001)	(5.970)	(2.538)	(4.618)
Unemp. rate (lag)	-0.077**	-0.091**	-0.139	-0.103*	-0.056
	(0.037)	(0.044)	(0.120)	(0.054)	(0.099)
Terrorim at dest., lag	0.007	0.003	-0.003	0.001	0.017
	(0.005)	(0.007)	(0.045)	(0.008)	(0.019)
Pseudo-R2	0.378	0.409	0.421	0.401	0.328
Origin-Year FE	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes
Observations	4600	4256	1837	3438	1234

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Heteroskedasticity-robust standard-errors in parenthesis. Variable Mig. is the ratio of movers from i to j over stayers derived from the standard RUM model. LS, MS, and HS correspond respectively to low-skilled (up to 8 years of education), middle-skilled (9-15 years of education), and high-skilled (4 years of education beyond high school). HSMS contains HS and MS individuals. Variable Eurobaro. is the share of individuals in the destination country that think immigration is one of the main two issues in their country. Control variables are the first stage residuals, log of GDP per capita, the unemployment rate, and the number of victims of terrorist attacks in the destination country lagged by one year. The development level variable is taken from the World Bank classification, with lower-middle income and upper-middle income countries grouped as 'Middle'. The reference level for interactions with development level is 'Low'. Non-parametric bootstrap was applied on both steps of the control function, using the full sample clustered by origin-destination, with replacement (1000 replications).

Table 1.B.3: Result with interaction on whether origin is in Europe

	(1) Mig	(2) Mig (HSMS)	(3) Mig (HS)	(4) Mig (MS)	(5) Mig (LS)
Attitudes, lag	-0.100***	-0.125***	-0.156***	-0.153***	-0.161***
	(0.012)	(0.013)	(0.040)	(0.016)	(0.027)
Attitudes, lag × Europe	0.002	0.017	0.008	0.023	-0.040
	(0.016)	(0.016)	(0.043)	(0.019)	(0.047)
1st stage resid. (lag)	0.118***	0.137***	0.155***	0.172***	0.184***
	(0.005)	(0.006)	(0.011)	(0.006)	(0.011)
GDP pc (log, lag)	1.206	1.744	4.675	1.519	2.996
	(1.835)	(2.028)	(5.900)	(2.583)	(4.555)
Unemp. rate (lag)	-0.085**	-0.102**	-0.192	-0.110**	-0.067
	(0.037)	(0.044)	(0.118)	(0.054)	(0.098)
Terrorim at dest., lag	0.008	0.003	-0.003	0.001	0.017
	(0.005)	(0.007)	(0.043)	(0.008)	(0.019)
Pseudo-R2	0.376	0.409	0.420	0.401	0.325
Origin-Year FE	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes
Observations	4651	4307	1851	3474	1234

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Heteroskedasticity-robust standard-errors in parenthesis. Variable Mig. is the ratio of movers from i to j over stayers derived from the standard RUM model. LS, MS, and HS correspond respectively to low-skilled (up to 8 years of education), middle-skilled (9-15 years of education), and high-skilled (4 years of education beyond high school). HSMS contains HS and MS individuals. Variable Eurobaro. is the share of individuals in the destination country that think immigration is one of the main two issues in their country. Control variables are the first stage residuals, log of GDP per capita, the unemployment rate, and the number of victims of terrorist attacks in the destination country lagged by one year. Europe dummy is equal to 1 if the origin country is in Europe, and 0 otherwise. Non-parametric bootstrap was applied on both steps of the control function, using the full sample clustered by origin-destination, with replacement (1000 replications).

Table 1.B.4: Result by skill using alternative instrument: victims outside adjacent countries

	(1) Mig (HSMS)	(2) Mig (HS)	(3) Mig (MS)	(4) Mig (LS)
Eurobaro., lag	-0.198***	-0.216***	-0.246***	-0.339***
	(0.011)	(0.030)	(0.013)	(0.026)
1st stage resid. (lag)	0.216***	0.217***	0.272***	0.348***
	(0.006)	(0.011)	(0.006)	(0.011)
GDP pc (log, lag)	1.434	4.456	1.136	2.089
	(2.003)	(5.883)	(2.547)	(4.590)
Unemp. rate (lag)	-0.146***	-0.226*	-0.165***	-0.165*
	(0.043)	(0.118)	(0.053)	(0.098)
Terrorim at dest., lag	0.001	-0.004	-0.000	0.015
	(0.007)	(0.044)	(0.008)	(0.018)
Pseudo-R2	0.409	0.420	0.401	0.325
Origin-Year FE	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes
Observations	4307	1851	3474	1234

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Heteroskedasticity-robust standard-errors in parenthesis. Variable Mig. is the ratio of movers from i to j over stayers derived from the standard RUM model. Variable Eurobaro. is the share of individuals in the destination country that think immigration is one of the main two issues in their country. LS, MS, and HS correspond respectively to low-skilled (up to 8 years of education), middle-skilled (9-15 years of education), and high-skilled (4 years of education beyond high school). HSMS contains HS and MS individuals. The instrument is the number of victims in terrorist attacks outside of the country and its adjacent countries. Control variables are the first stage residuals, log of GDP per capita, the unemployment rate, and the number of victims of terrorist attacks in the destination country lagged by one year. Non-parametric bootstrap was applied on both steps of the control function, using the full sample clustered by origin-destination, with replacement (1000 replications).

Table 1.B.5: Result by skill using alternative instrument: victims outside destination country

	(1) Mig (HSMS)	(2) Mig (HS)	(3) Mig (MS)	(4) Mig (LS)
Eurobaro., lag	-0.197***	-0.199***	-0.244***	-0.321***
	(0.011)	(0.030)	(0.013)	(0.026)
1st stage resid. (lag)	0.215***	0.199***	0.270***	0.329***
	(0.006)	(0.011)	(0.006)	(0.011)
GDP pc (log, lag)	1.455	4.599	1.164	2.213
	(2.003)	(5.883)	(2.547)	(4.590)
Unemp. rate (lag)	-0.145***	-0.215*	-0.164***	-0.154
	(0.043)	(0.118)	(0.053)	(0.098)
Terrorim at dest., lag	0.001	-0.004	-0.001	0.016
	(0.007)	(0.044)	(0.008)	(0.018)
Pseudo-R2	0.409	0.420	0.401	0.325
Origin-Year FE	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes
Observations	4307	1851	3474	1234

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Heteroskedasticity-robust standard-errors in parenthesis. Variable Mig. is the ratio of movers from i to j over stayers derived from the standard RUM model. LS, MS, and HS correspond respectively to low-skilled (up to 8 years of education), middle-skilled (9-15 years of education), and high-skilled (4 years of education beyond high school). HSMS contains HS and MS individuals. Variable Eurobaro. is the share of individuals in the destination country that think immigration is one of the main two issues in their country. The instrument is the number of victims in terrorist attacks outside of the country. Control variables are the first stage residuals, log of GDP per capita, the unemployment rate, and the number of victims of terrorist attacks in the destination country lagged by one year. Non-parametric bootstrap was applied on both steps of the control function, using the full sample clustered by origin-destination, with replacement (1000 replications).

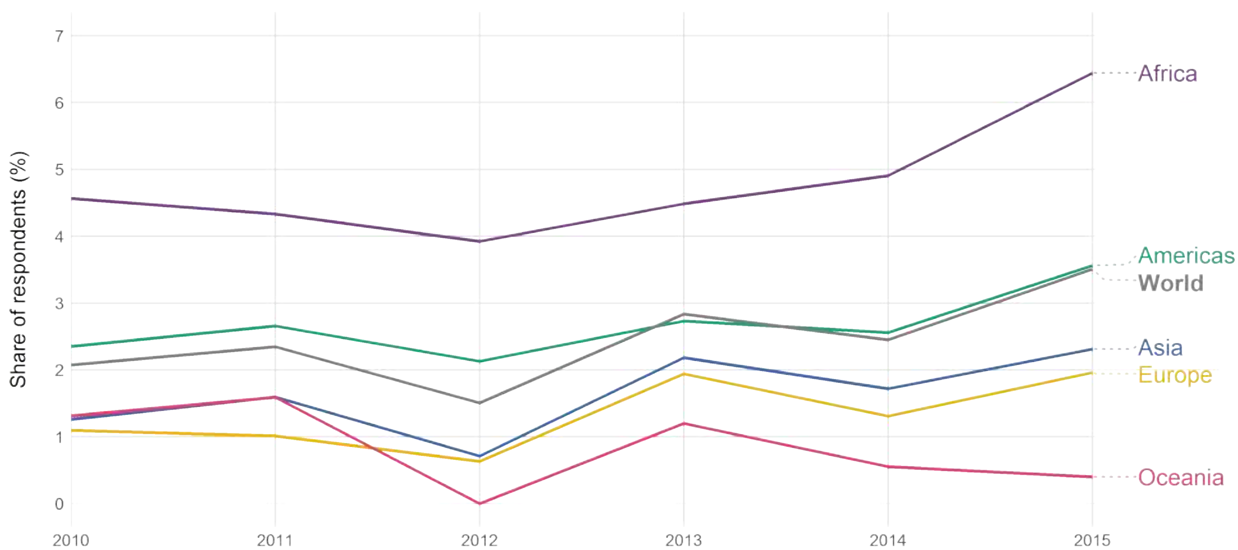


Figure 1.B.1: Share of respondents with a plan to migrate permanently in the 12 next months

References

Alesina, A. and M. Tabellini (2024). "The Political Effects of Immigration: Culture or Economics?" In: *Journal of Economic Literature* 62(1), pp. 5–46.

Arango, J. (2000). "Explaining Migration: A Critical View". In: *International Social Science Journal* 52(165), pp. 283–296.

Beine, M., S. Bertoli, et al. (2016). "A Practitioners' Guide to Gravity Models of International Migration". In: *The World Economy* 39(4), pp. 496–512.

Beine, M., M. Bierlaire, et al. (2024). *The Impact of a Possible Trump Reelection on Mexican Immigration Pressures in Alternative Countries*. CESifo Working Paper 11195.

Beine, M., A. Dupuy, et al. (2020). *Migration Intentions: Data from a Field Study in Albania*.

Bertoli, S. and I. Ruyssen (2018). "Networks and Migrants' Intended Destination". In: *Journal of Economic Geography* 18, pp. 705–728.

Böhmelt, T., V. Bove, et al. (2020). "Can Terrorism Abroad Influence Migration Attitudes at Home?" In: *American Journal of Political Science* 64(3), pp. 437–451.

Bove, V., T. Böhmelt, et al. (2021). "Terrorism Abroad and Migration Policies at Home". In: *Journal of European Public Policy* 28(2), pp. 190–207.

Card, D., C. Dustmann, et al. (2012). "Immigration, Wages, and Compositional Amenities". In: *Journal of the European Economic Association* 10(1), pp. 78–119.

Clemens, M. A. and M. Mendola (2020). *Migration from Developing Countries: Selection, Income Elasticity and Simpson's Paradox*. IZA Discussion Paper Series 13612.

Correia, S., P. Guimarães, et al. (2020). "Fast Poisson Estimation with High-Dimensional Fixed Effects". In: *The Stata Journal* 20(1), pp. 95–115.

Cruz, E., S. J. D'Alessio, et al. (2020). "Decisions Made in Terror: Testing the Relationship between Terrorism and Immigration". In: *Migration Studies* 8(4), pp. 573–588.

Docquier, F., G. Peri, et al. (2014). "The Cross-country Determinants of Potential and Actual Migration". In: *International Migration Review* 48(1), pp. 37–99.

Enders, W., T. Sandler, et al. (2011). "Domestic versus Transnational Terrorism: Data, Decomposition, and Dynamics". In: *Journal of Peace Research* 48(3), pp. 319–337.

Ermisch, J. and C. H. Mulder (2019). "Migration Versus Immobility, and Ties to Parents". In: *European Journal of Population* 35(3), pp. 587–608.

Facchini, G. and A. M. Mayda (2012). "Individual Attitudes Towards Skilled Migration: An Empirical Analysis Across Countries". In: *The World Economy* 35(2), pp. 183–196.

Ferrín, M., M. Mancosu, et al. (2020). "Terrorist Attacks and Europeans' Attitudes towards Immigrants: An Experimental Approach". In: *European Journal of Political Research* 59(3), pp. 491–516.

Foubert, K. and I. Ruyssen (2024). "Global Migration and the Role of Terrorist Attacks". In: *Journal of Economic Behavior & Organization* 220, pp. 507–530.

Gorinas, C. and M. Pytliková (2017). “The Influence of Attitudes toward Immigrants on International Migration”. In: *International Migration Review* 51(2), pp. 416–451.

Hatton, T. J. (2021). “Public Opinion on Immigration in Europe: Preference and Salience”. In: *European Journal of Political Economy* 66, p. 101969.

Hatton, T. J. and A. Leigh (2011). “Immigrants Assimilate as Communities, Not Just as Individuals”. In: *Journal of Population Economics* 24(2), pp. 389–419.

Head, K. and T. Mayer (2014). “Gravity Equations: Workhorse,Toolkit, and Cookbook”. In: *Handbook of International Economics*. Vol. 4. Elsevier, pp. 131–195.

Helbling, M. and D. Meierrieks (2020). “Terrorism and Migration: An Overview”. In: *British Journal of Political Science*, pp. 1–20.

Iasio, V. D. and J. Wahba (2023). *Natives' Attitudes and Immigration Flows to Europe*, p. 38.

Legewie, J. (2013). “Terrorist Events and Attitudes toward Immigrants: A Natural Experiment”. In: *American Journal of Sociology* 118(5), pp. 1199–1245.

Lin, W. and J. M. Wooldridge (2019). “Testing and Correcting for Endogeneity in Nonlinear Unobserved Effects Models”. In: *Panel Data Econometrics*. Elsevier, pp. 21–43.

Manchin, M. and S. Orazbayev (2018). “Social Networks and the Intention to Migrate”. In: *World Development* 109, pp. 360–374.

Mincer, J. (1978). “Family Migration Decisions”. In: *Journal of Political Economy* 86(5), pp. 749–773. JSTOR: 1828408.

Miroudot, S. and D. Rigo (2021). “Multinational Production and Investment Provisions in Preferential Trade Agreements”. In: *Journal of Economic Geography*, pp. 1–34.

Nussio, E., V. Bove, et al. (2019). “The Consequences of Terrorism on Migration Attitudes across Europe”. In: *Political Geography* 75, p. 102047.

Ortega, F. and G. Peri (2013). “The Effect of Income and Immigration Policies on International Migration”. In: *Migration Studies* 1(1), pp. 47–74.

Ritchey, P. N. (1976). “Explanations of Migration”. In: *Annual Review of Sociology* 2, pp. 363–404. JSTOR: 2946097.

Santos Silva, J. and S. Tenreyro (2006). “The Log of Gravity”. In: *The Review of Economics and Statistics* 88(4), pp. 641–658.

Santos Silva, J. and S. Tenreyro (2011). “Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator”. In: *Economics Letters* 112(2), pp. 220–222.

Schewel, K. (2020). “Understanding Immobility: Moving Beyond the Mobility Bias in Migration Studies”. In: *International Migration Review* 54(2), pp. 328–355.

Slotwinski, M. and A. Stutzer (2019). “The Deterrent Effect of an Anti-Minaret Vote on Foreigners' Location Choices”. In: *Journal of Population Economics* 32(3), pp. 1043–1095.

Wooldridge, J. M. (2014). “Quasi-Maximum Likelihood Estimation and Testing for Nonlinear Models with Endogenous Explanatory Variables”. In: *Journal of Econometrics* 182(1), pp. 226–234.

Wooldridge, J. M. (2015). “Control Function Methods in Applied Econometrics”. In: *Journal of Human Resources* 50(2), pp. 420–445.

World Bank (n.d.). *World Development Indicators*.

Chapter 2

Populist leaders and international migration

Abstract

While immigration is often considered as an important cause of populism, the consequences of populist leaders on international migration have received much less attention. In this paper, I explore the dynamic effects of populist leaders leaders on both immigration and emigration, how they affect the selectivity of immigration, as well as the underlying mechanisms. I use data on populist leaders and migration flows since 1960 and the generalized synthetic control method for identification. I find that having a populist leader significantly decreases the growth rate of immigration, in particular for low-skilled immigration. However, there is no effect on emigration growth rates. I use rich datasets on migration policies and migration intentions to explore potential mechanisms, and I find that right-wing populist leaders implement more restrictive migration policies. In addition to reducing inflows, those policies also increase the intentions of low-skilled migrants who were already settled in the country to leave and go back to their home country. On the other hand, having a populist leader doesn't change the attractiveness of the country in the eyes of potential migrants.

2.1 Introduction

Populism has been flourishing in recent years, both in developed and emergent countries. Classic examples such as Trump's election and Brexit in 2016 are only the tip of the iceberg. It is now frequent to see populist parties win elections or obtain an important share of votes that put them in a position to heavily influence national and international laws.

A large literature has grown in the last decade to explain the causes of this rise of populism. Among them, the role of immigration on populism and political attitudes has been largely documented across countries (Otto and Steinhardt, 2014; Mendez and Cutillas, 2014; Barone et al., 2016; Halla et al., 2017; Mayda et al., 2022; Chen, 2023). Many studies have focused on the effect of immigration on attitudes towards migrants and have made clear that the skill composition of migration flows is key in the formation of these attitudes (Hainmueller and Hiscox, 2010; Facchini and Mayda, 2012; Hainmueller and Hopkins, 2015; Hopkins, 2015; Edo et al., 2019; Moriconi et al., 2022; Mayda et al., 2022). Recently, Docquier et al. (2023) have extended this investigation to the last 60 years in developed countries, focusing on the effect of the skill composition of immigration and trade on populism. They find that low-skilled immigration does not affect the share of votes for populist parties but rather transfers votes from left-wing populist parties to right-wing ones. However, while there are numerous studies on the effect of immigration on political attitudes, the potential reverse relationship is much less explored.

In this paper, I explore the causal effect of having a populist leader on immigration and emigration flows in the subsequent years. To do so, I rely on a dataset created by Funke et al. (2023) that classifies every leader as populist or non-populist in more than 60 countries since 1960. In the absence of credible counterfactuals for countries with populist leaders, I use the synthetic control method to build a synthetic country that matches the treated country in the decade before the populist arrives at power. This should mitigate the endogeneity issue as it ensures that the counterfactual matches the trend of the outcome and other controls of the treated country before treatment. Since I focus on more than a dozen settings, I use the generalized synthetic control (Xu, 2017) to take into account staggered adoption and variable treatment length.

I find that having a populist leader leads to a significant decrease in the growth rate of immigration by more than 20 percentage points during the ruling period but that this growth rate comes back to its original level after the populist leader leaves power. This negative effect is stronger for low-skilled immigration flows than for high-skilled ones. I also show that the size of this effect varies largely across countries, which motivates the pursuit of country-specific studies. However, I don't find any significant effect on the growth rate of emigration. This contrasts with the results of Docquier and Vasilakis (2023) who find that an increase in right-populism leads to more high-skilled emigration.

In a second step, I try to disentangle the mechanisms behind this reduction in immigration between factors in the destination country (where the populist leader arrives at power) and factors in the origin countries. Using data from the DEMIG policy project I find that right-wing populist leaders implement more restrictive immigration policies targeting specifically low-skilled migrants, and that left-wing populist leaders don't increase the number of migration policies. Finally, using data from the Gallup World Polls, I don't find any clear evidence for a decrease in attractiveness for countries that have a populist leaders. This suggests that migration policies are the main mechanism through which populist leaders affect immigration flows.

This paper contributes to the existing literature on migration and populism. The main challenge in this literature is to overcome the endogeneity arising from omitted variables that affect both migration flows and the evolution of populism, such as the economic conditions. This is the case both for internal and international migration. Most papers that look at the effect of populism on migration address this issue by using regression discontinuity designs in close elections settings at the municipality level. Using the Italian context in the period 2002–2014, Bracco et al. (2018) find that having a Lega Nord mayor has a negative effect on immigrants flows but has no effect on the decision to leave of migrants who already settled there. In a comparable setting, Cerqua and Zampollo (2023) find a similar result on immigration flows and explain it by the increase in inhospitality in municipalities lead by an anti-immigration coalition, rather than by actual policies. Focusing on France, Schmutz and Verdugo (2023) find that electing a left-wing mayor increases the growth rate of the immigrant population at the municipality level.

Bellodi et al. (2024) take an another approach by using a combination of collective memory and trigger variables to extract some exogenous variation in populist vote shares. They find that having a populist mayor reduces the attractiveness of the municipality, leading to outflows of high-skilled natives who resettle in other Italian municipalities. An important result of their paper is also that there is virtually no effect on foreigners.

Those papers study mostly the internal mobility of foreign-born and natives. Only two papers focus on international migration. Pan (2023) uses a gravity model to study the effect of populism on the international movements of inventors and finds that it discourages them from coming. In particular, right-wing populism always decreases inventors immigration while left-wing populism has mixed effects, depending on whether a populist is also at power in the origin country. Demirci (2023) uses the synthetic control method to explore the effect of populism on emigration of youth. Using the rise of populist leaders in four countries in the recent years (Hungary, Ukraine, Venezuela, and Indonesia), he finds that students are more likely to leave the country and study abroad, and that this would most likely come from higher desires to live abroad permanently rather than from a decline in the quality of education in the origin country. In comparison to those papers, my work focuses on international immigration and emigration flows and takes all types of migrants into account.

The rest of the paper is organized as follows. In section 2.2, I present the variants of the synthetic control method that are used in the paper. Section 2.3 presents the sources of the data on populism and migration. I describe the results in section 2.4 and explain the potential mechanisms in section 2.5. Section 2.6 concludes.

2.2 Method

I investigate the effect of populism on immigration. One obvious issue is that it is hard, if not impossible, to obtain a credible counterfactual for the treated country. Indeed, there can be large structural differences in the economic and political landscapes of industrialized or emerging countries that affect migration flows towards these countries[18]. Therefore, comparing each treated country to another country in the control group would lead to unreliable results.

To overcome these issues, I use the synthetic control method (SCM) to create an artificial counterfactual that will be used to see the effect of populism on immigration. The aim of the SCM is to use an optimization algorithm to find the weighted combination of untreated countries that generates the most accurate fit for the trend of the treated country in the pre-treatment period. This synthetic "doppelganger" would therefore correctly match the structural characteristics of the treated country. Using the notation of Abadie (2021), consider a pool of $J+1$ units where the first unit $j=1$ is the treated unit and the other $j=2,\ldots,J+1$ are the control units (also named "donor pool"). I have T periods and the treatment occurs in T_0. For each unit, I also have k predictors of the outcome. I look for the effect of having a populist leader, which is $\tau_{1t} = Y_{1t}^I - Y_{1t}^N$ where Y_{1t}^I is the outcome of the treated unit and Y_{1t}^N is the outcome that this unit would have had without the populist leader for $t > T_0$. Therefore, I look for a reliable counterfactual that would allow me to get Y_{1t}^N.

To be credible, a counterfactual needs to correctly represent the trajectory the treated country would have had without the populist leader. To do so, the algorithm finds the optimal

[18] This is a less important issue for papers exploring the within-country variation in populism since regions or municipalities share several common features such as the language or their history. Of course, there can also be important differences between regions in a country, but those are usually less important than differences between two countries separated by thousands of kilometers.

weights $W^* = (w_2^*, \dots, w_{J+1}^*)$ that minimize:

$$||\mathbf{X_1} - \mathbf{X_0}\mathbf{W}|| = \left(\sum_{h=1}^{k} v_h (X_{h1} - w_2 X_{h2} - \dots - w_{J+1} X_{hJ+1})^2\right)^{1/2} \tag{2.1}$$

with the restriction that unit weights W^* must be between 0 and 1 and their sum must be equal to 1. Here, v_h for $h = 1, \dots, k$ is the importance given to each predictor in the construction of the synthetic control. The Xs are the predictors used to build the synthetic control. Here, I use the past values of the outcome Y and of the GDP per capita for each year in the pre-treatment period to create the synthetic control. I can then use $\mathbf{W}^*$ to compute synthetic control and get the estimated treatment effect:

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \tag{2.2}$$

Therefore, in the case of a single event, I can determine the effect of populism on immigration as the difference between the trends of the synthetic control and the treated country in the post-treatment period.

However, an additional challenge in my setting is that I want to determine the effect of populism on immigration based on multiple events. This introduces several difficulties. First, there is staggered treatment because populist leaders arrive at power in different years between 1960 and 2020. Second, contrarily to more classic staggered adoption settings where a treated unit stays treated until the end of the period of interest, a country is generally treated only temporarily because populist leaders come and go. Therefore, one country can be treated, then untreated for a few years or decades, and then treated again.

To address those challenges, I use the generalized synthetic control method (Xu, 2017). This variant of the synthetic control method is a mixture of the SCM and of the difference-in-differences method, as it includes a combination of unit and time fixed effects when fitting the synthetic controls. An attractive feature of this method is that it allows for multiple treated units and variable treatment periods. Moreover, it provides confidence intervals via a parametric bootstrap procedure. I use the R implementation in the package `gsynth` (Xu and Liu, 2022). To assess the robustness of the generalized SCM, I also run estimations with time placebos by setting the treatment period in t_{-2}[19].

[19] The synthetic control method requires as many pre-treatment periods as possible to be correctly fit. Given that I have 10 years in the pre-treatment period for regular estimations, I choose to remove two years only when setting the time placebo.

To show that the results do not depend solely on the method used, I use several variants of the SCM recently developed, namely the partially pooled SCM (Ben-Michael et al., 2021) and the synthetic difference-in-differences (SDiD, Arkhangelsky et al., 2021). Those variants are detailed further in appendix 2.B. Although they do not provide confidence intervals, the pre-treatment fit and the effect in the post-treatment period are extremely similar to those produced by the generalized synthetic control.

2.3 Data

2.3.1 Populism

The notion of ”populism” is hard to define. Guriev and Papaioannou (2022) provide an extensive literature review on this question and conclude that in the current state, there is not one definition of populism but rather many partially overlapping ones. Still, some features are shared across all definitions. The first one is *anti-elitism*, the way that populist leaders have of presenting themselves as the true representatives and protectors of the people against the disconnected establishment. The second one is *anti-pluralism*, which is the willingness to let minorities aside and prevent differences in opinions[20]. This absence of clear consensus on the definition of populism also means that measuring it is subject to interpretation. Some have used political manifestos to study politicians' discourse, while others have relied on the opinions of experts for each country. Some of these measures are continuous, and others are categorical.

In this paper, I use a recent dataset created by Funke et al. (2023) in which each country leader (Prime Minister or President) is classified as either populist or not populist. This dataset covers 60 countries in the period 1900-2018. This classification is based on a literature review on each leader and only looks at the discourse of the politicians. Specifically, a leader is classified as populist if “he or she divides society into two artificial groups – ’the people’ vs. ’the elites’ – and then claims to be the sole representative of the true people” (Funke et al., 2023). It is important to note that the classification of a leader does *not* depend on the policies they implemented.

The fact that this classification is binary is both one of its strengths and a limitation. Indeed, this classification allows me to use event-study methods which would not be available with a continuous measure. On the other hand, political parties and leaders may not be classified as populists while they share a lot of common features with leaders that are classified as populists. Using this binary classification means that we can't distinguish the ”intensity of

[20] These two features were already part of the definition of Mudde (2004).

	Mean	SD	Min	Median	Max
Δ total immig.	5.56	24.15	-94.26	3.40	566.67
Δ HS immig.	6.42	23.62	-100.00	3.94	404.91
Δ LS immig.	5.48	25.45	-94.00	3.09	629.42

Table 2.3.1: Summary statistics of the outcome (growth rate of immigration in total and in several subgroups, in %)[21].

populism".

2.3.2 Migration

Studies focusing on migration at a macro level in a panel setting have to deal with very incomplete data. The large differences in data collection or in definitions of who is a migrant make it hard to harmonize immigration numbers. Even when the data is harmonized, like in the DEMIG dataset, it still suffers from many missing values meaning that we need to rely extensively on imputation methods. This issue is even more salient when one wants to cover several decades as here. Finally, having reliable data is even harder when one is interested in bilateral migration flows, which is something that is used here to see whether there are heterogeneous effects depending on migrants' country of origin.

Given these constraints, I rely on a dataset recently created by Standaert and Rayp (2022), who use a Bayesian state-space model to compute yearly bilateral migration flows. This combines information on migrants stocks, gross and net migration flows to a country, birth, and deaths to produce a single estimate of the size of migration flows between country pairs. They find that this data imputation performs better in terms of correlation with other existing measures than other kind of imputations. This dataset goes from 1960 to 2020 and covers almost all possible origin-destination pairs. The fact that this data has a yearly dimension is essential for me since the synthetic control is more reliable when there is a large number of observations in pre-treatment period (Abadie, 2021). Moreover, having bilateral flows allows me to focus more specifically on some groups of origin countries.

Still, an aspect that is not covered by this dataset but essential to my analysis is the one of skill level. The question of skill selection is central in migration economics. However, measuring the skill level of migrants is an arduous task. I use data on the skill distribution in 1990 and 2000 coming from Artuc et al. (2015) and Arslan et al. (2015) respectively. This provides the share of high-skilled (i.e. college-educated) migrants coming from a specific origin in a specific destination in 1990 and in 2000. I use this bilateral share in the other years,

[21] The minimum value for several immigration variables is -100 because it is what happens when we pass from a positive number of migrants to 0 migrants from one year to the other ($(0 - x)/x \times 100 = -100$).

therefore assuming that the skill distribution will stay the same across destinations.

Using the raw counts of migrants would make it difficult to apply the SCM since it computes a weighted combination that must fit the trend of the treated country. Therefore, I standardize my outcome by taking the growth rate of immigration and emigration per capita. There are sometimes large migration shocks that significantly affect the growth rate of immigration, so I take the moving average of immigration on a three-year window. Table 2.3.1 shows the summary statistics for the growth rate of immigration in different categories.

As an alternative to using the skill distribution, I also use a recent dataset on scholars migration over the period 1998-2020 (Akbaritabar et al., 2024). This dataset uses two main sources compiling scholars information, namely Scopus and OpenAlex. This dataset gives information on immigration and emigration flows of scholars in more than 200 countries. I compute the same measurements as for the data from Standaert and Rayp (2022) and I use this data to check the robustness of the effect on migration flows of high-skilled individuals.

2.3.3 Migration policies

The DEMIG policy dataset (DEMIG, 2015) covers 42 countries on the period 1945-2014[22]. It contains information on several aspects of migration policies in each of these countries: whether it increases or decreases the restrictiveness of migration policies, whether the change is done unilaterally or with other countries, the size of the change, details about the targeted group including countries of origin if specified, and several others. Overall, it provides information reviewed by national experts on more than 6,500 migration policy changes. I use this information to explore the effects that populist leaders have on the legal barriers implemented to reduce immigration.

2.3.4 Migrants' preferences

The Gallup World Polls (GWP) are annual surveys conducted in almost all countries in the world. They cover a vast amount of topics, and have been increasingly used in the migration literature in the last decade. Indeed, since 2007, respondents are asked whether they would like to migrate *permanently* to another country and, if so, the name of this country. Here, I use this question to analyze whether the arrival at power of a populist has an effect on the attractiveness of this country, i.e on the rank of this country when we compute the number of

[22] Some countries, such as Slovenia, are only covered since 1990 but the large majority of the country are covered at least since 1945.

respondents who wish to migrate to each destination[23]. I use this dataset to analyze whether the arrival at power of populist leader affected the attractiveness of the country in terms of number of Gallup respondents who want to migrate there.

2.3.5 Other

As explained in section 2.2, I also use the GDP per capita to fit the synthetic control. This data is obtained from the World Bank (World Bank, n.d.).

2.4 Results

2.4.1 On immigration

Figure 2.4.1 shows the main result of the paper. It displays the difference between the observed growth rate of immigration and the synthetic one. A negative value means that the observed growth rate of immigration is lower than what it would have been without the populist leader. The difference is statistically insignificant and close to zero in the pre-treatment period, indicating that the synthetic control fits the trend of the observed country relatively well. Looking at the post-treatment period, we can see that having a populist leader has a negative effect on the growth rate of immigration, and this effect is statistically significant after the second year. Its magnitude is also large: after 3 years, the growth rate of immigration is 18 percentage points lower than in the counterfactual scenario. Given that its mean value is 5.56 (cf table 2.3.1), this means that the number of immigrants arriving each year decreases.

Figure 2.4.1 also shows that this effect is more important, both in magnitude and in statistical significance, when we focus on inflows of low-skilled migrants. This would corroborate the intuition that populist leaders tend to target specifically this group of migrants when they implement or reinforce anti-immigration policies. Figure 2.C.1 in Appendix shows that setting the treatment time in t_{-2} doesn't change those results.

Using scholars' immigration as an alternative measure for high-skilled immigration leads to similar results. Figure 2.D.1 in the appendix shows that having a populist leader has no effect on the arrival of scholars in a country.

So far, those results are aggregated and don't allow us to see heterogeneity by event. Instead of aggregating all events on several time periods, we can also compute the ATT for each event, using the 10 years in the post-treatment period. Figure 2.4.2 suggests that there is

[23] For some waves, additional questions were asked to assess whether this is simply a desire to migrate or if the respondent already planned to migrate and started preparing for this. However, the number of respondents who started preparing is much smaller, so I only use the desire to migrate.

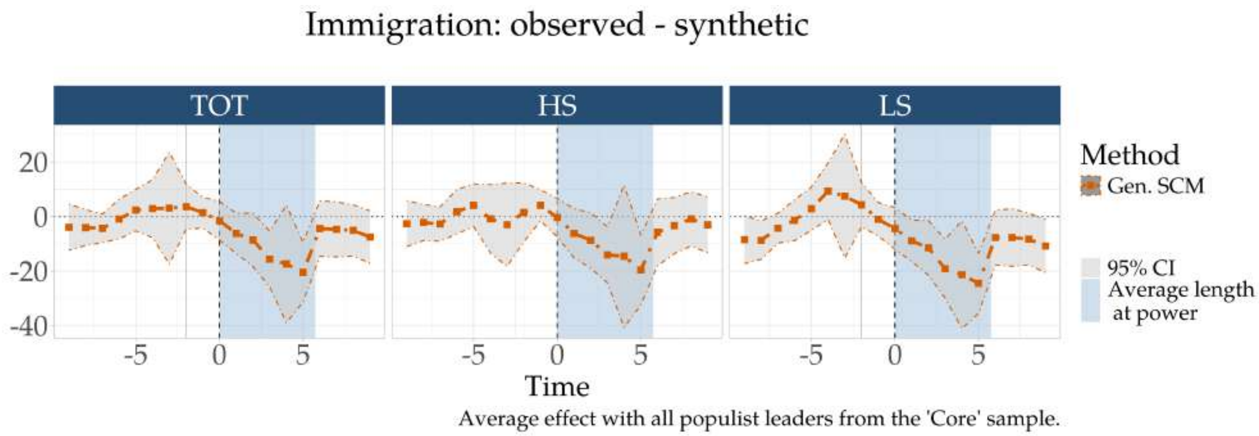


Figure 2.4.1: Effect of having a populist leader on growth rate of immigration (in percentage points). The orange dashed line shows the difference between the observed growth rate of immigration and the synthetic one. The gray band shows the 95% CI, computed with bootstrap (1,000 repetitions). The vertical dotted line indicates the period at which the populist leader arrives at power.

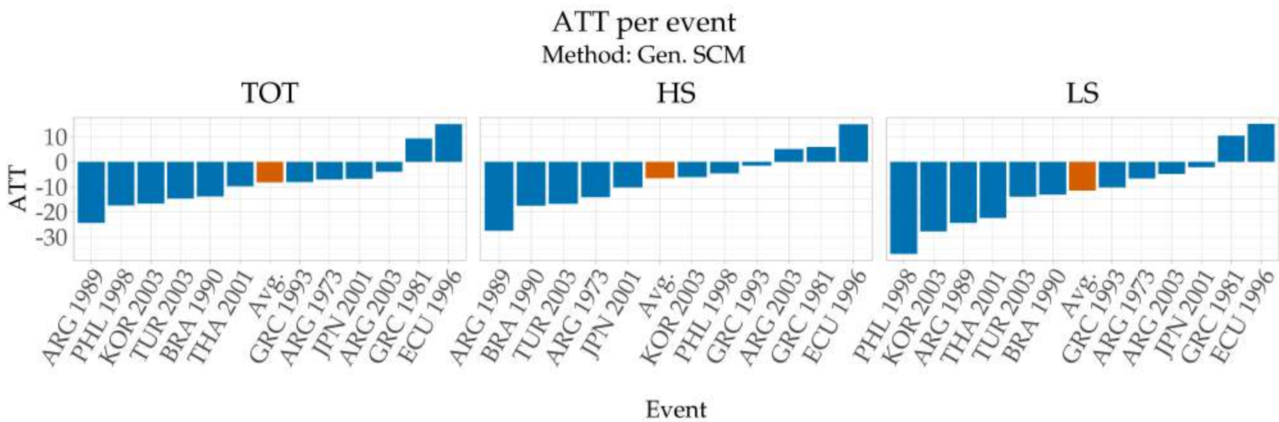


Figure 2.4.2: ATT of having a populist leader on growth rate of immigration (in percentage points). This is averaged by event over the full post-treatment period. The orange bar shows the full average.

a lot of variability across events. At the extreme, we see that the election of Abdalá Bucaram in Ecuador in 1996 had the opposite effect on immigration growth rates. It is interesting to see that the effect can be positive for several events. This shows that populism doesn't have the same effect everywhere and that it is worth exploring those effects in different countries and time periods.

2.4.2 On emigration

I now turn to the effect of populism on emigration growth rates. While Docquier and Vasilakis (2023) find an increase in outflows of high-skilled individuals following the arrival of a populist leader, figure 2.4.3 suggests that low-skilled individuals tend to emigrate more instead. However, those results should be subject to caution as there is only a single point in

Emigration: observed - synthetic

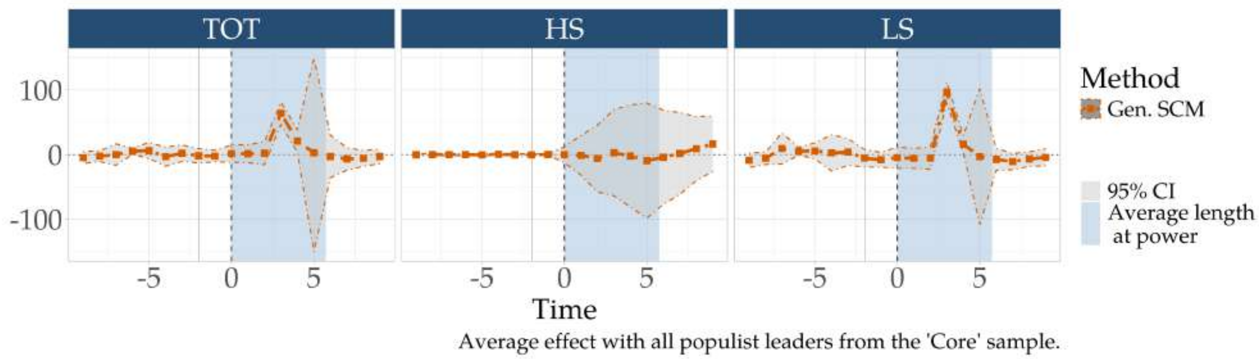


Figure 2.4.3: Effect of having a populist leader on growth rate of emigration (in percentage points).

the post-treatment period that is statistically significant.

Moreover, splitting this aggregate result by event in figure 2.4.4 gives several insights. First, the positive effect that we observe on total emigration and low-skilled emigration is mostly driven by the case of Ecuador in 1996, which is once again an outlier in the distribution of ATT by event. Indeed, at the end of the 1990s, Ecuador saw an impressive increase in emigration following its largest economic crisis in the 20th century (Jokisch and Pribilsky, 2002). Figures 2.E.2 in the appendix shows the results when I remove Ecuador-1996 from the list of events. The aggregate effect on emigration becomes insignificant, no matter the skill level, indicating that this event is driving most of the result on emigration. This is also the case when I focus on scholars emigration flows, as shown by figure 2.D.2 in the appendix.

Second, although the aggregate effect on high-skilled emigration is statistically insignificant and its magnitude is close to 0, there is a lot of heterogeneity in the ATT by event, and several populist leaders led to a significant increase in high-skilled emigration, such as in Turkey in 2003.

2.4.3 Left-wing and right-wing populism

I now separate populist leaders between left-wing and right-wing ones. Docquier et al. (2023) find that low-skilled immigration is mostly associated with right-wing populism and not so much with left-wing populism. Similarly, Pan (2023) finds that inventors, who mostly fall in the standard definition of ”high-skilled” individuals, mostly react to the arrival at power of a right-wing leader. Therefore, I would expect the negative effect of populism on immigration to be driven by right-wing leaders. Figures 2.4.5 and 2.4.6 show the baseline results when we only keep left-wing and right-wing leaders respectively. I see that the overall effect on immigration is driven by right-wing populist leaders. However, when we split migrants

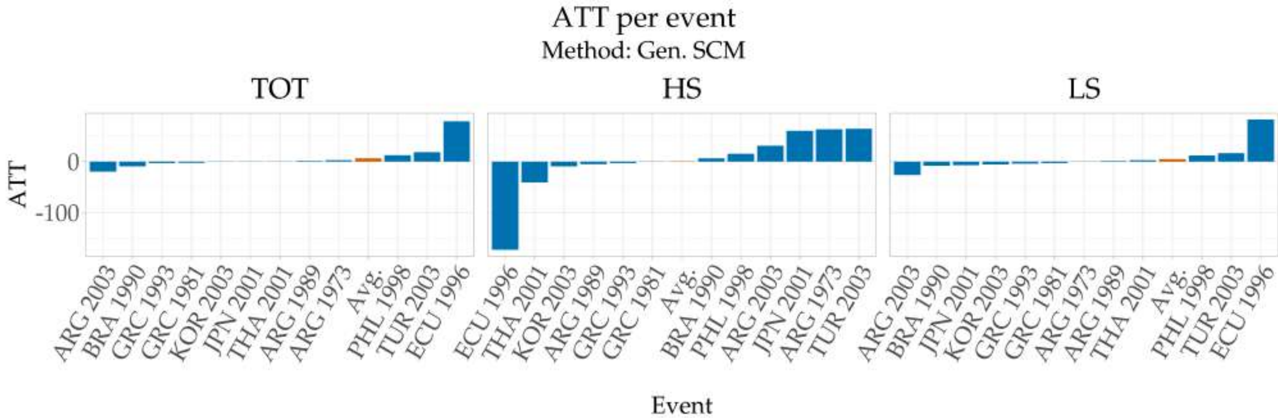


Figure 2.4.4: ATT of having a populist leader on growth rate of emigration (in percentage points). This is averaged by event over the full post-treatment period. The orange bar shows the full average.

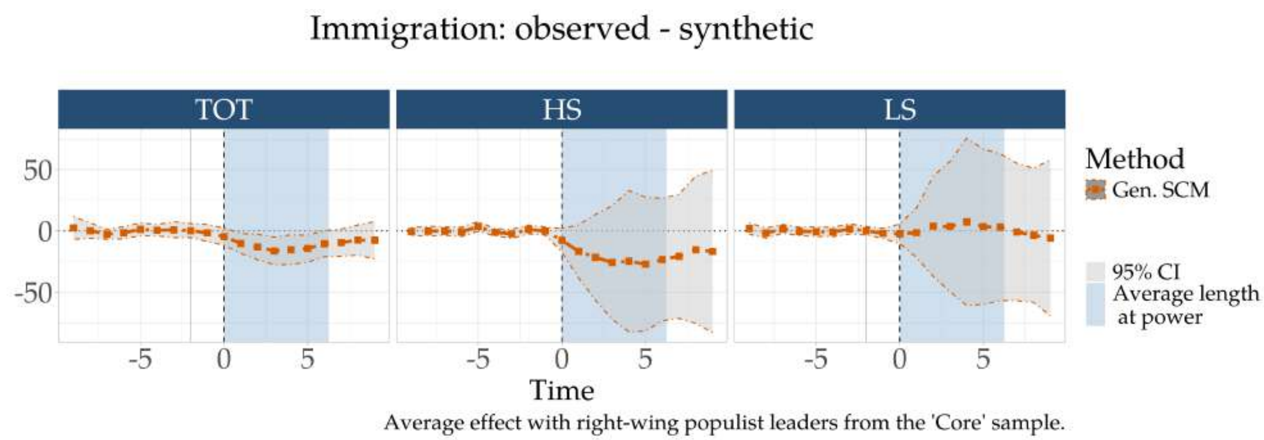


Figure 2.4.5: Effect of having a right-wing populist leader on growth rate of immigration (in percentage points).

by the skill level, then confidence intervals become extremely large and results are mostly statistically insignificant, whether I focus on left-wing or right-wing populist leaders.

2.5 Mechanisms

My results suggest so far that having a populist leader reduces immigration in the following years, but they do not allow me to know where this effect comes from. In this section, I explore two types of mechanisms. Firstly, I focus on the destination country, i.e the country that has a populist leader. Using the DEMIG policy dataset, I analyze whether this decrease in immigration is due to the implementation of more restrictive immigration policies. Secondly, I focus on the migrants, or rather the intended migrants. Using the Gallup World Polls, I explore whether the arrival of a populist at power in the destination country has an effect on the number of people who have a desire to migrate to this country.

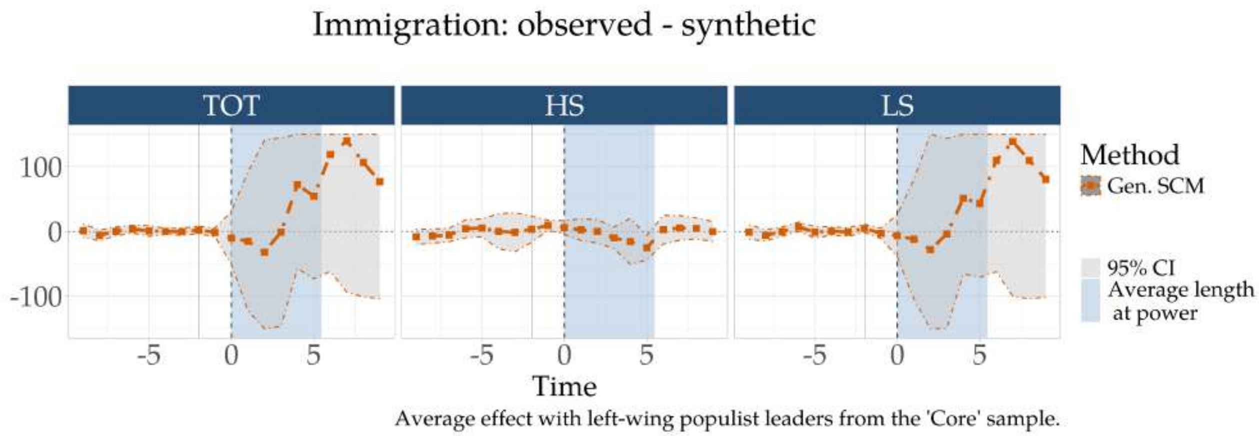


Figure 2.4.6: Effect of having a left-wing populist leader on growth rate of immigration (in percentage points).

2.5.1 Migration policies

I combine data from the DEMIG policy project with the data on populist leaders to analyze whether policies implemented by populist leaders can explain my results. The objective is to compare the number of migration policies when a populist or a non-populist is at power. To do so, I use a simple OLS specification in which I predict the number of migration policies by an indicator for populism:

$$Y_{it} = \alpha + \beta Pop_{it} + \nu_i + \tau_t + \varepsilon_{jt} \tag{2.3}$$

Y_{it} the number of migration policies in country i at year t. This can be the count of all migration policies, or only restrictive or only permissive migration policies. Pop_{it} is a dummy variable that is equal to 1 if a populist leader is at power in country i at year t. I consider separately all populist leaders, only left-wing, and only right-wing populist leaders. Overall, I have 3×3 estimated coefficients. Finally, the equation above also comprises time fixed effects τ_t and country fixed effects ν_i to control for global trends and country-specific time-invariant factors.

Table 2.5.1 shows the results of the 9 estimations. Overall, we can see that the only case in which having a populist leader is significantly associated to migration policies is for right-wing populists. In particular, there is a positive relationship between having a right-wing populist leader at power and the number of restrictive migration policies implemented in a given year. Interestingly, I find no effect of left-wing populist leaders, no matter the type of migration policy.

So far we considered all restrictive migration policies without taking into account the

	All mig. pol.	Restrictive mig. pol.	Permissive mig. pol.
All pop. leaders			
Populism	0.017	0.156	-0.096
	(0.124)	(0.174)	(0.138)
RW pop. leaders			
Populism	0.103	0.374**	-0.118
	(0.128)	(0.173)	(0.154)
LW pop. leaders			
Populism	-0.171	-0.469	-0.008
	(0.257)	(0.378)	(0.292)
Year FE	Yes	Yes	Yes
Country FE	Yes	Yes	Yes
Obs.	870	773	870

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Heteroskedasticity-robust standard errors in parenthesis. The dependent variable is the number of migration policies, the number of more restrictive migration policies only, and the number of more permissive migration policies only. The interest variable is an indicator of populism, considering all populist leaders in the first row, right-wing populist leaders only in the second row, and left-wing populist leaders only in the third row.

Table 2.5.1: Effect of populism on the number of migration policies.

importance of the change. However, changes implemented by right-wing populists could be minor and apply to only a small fraction of migrants groups. The DEMIG policy dataset provides an indicator on the magnitude of the change. This indicator can take the following values: ”fine-tuning change”, ”minor change”, ”mid-level change”, ”major change”[24]. I run the same regressions as above but considering only ”major” policy changes, i.e “changes that affect an entire migrant category and introduce or remove a new policy instrument” (DEMIG, 2015). As shown in table 2.5.2, the results are very similar: right-wing populist leaders are associated with a higher number of major restrictive migration policies, and we don’t see this kind of relationship for other combinations of populism and output.

2.5.2 Migrants’ preferences

Seeking a causal effect of the arrival at power of a populist on the attractiveness of this country is challenging. Multiple variables can have an effect both on the rise of populism and on the attractiveness of a country, which leads to endogeneity issues. In addition to that, taking into account diversion effects towards alternative destinations also proves to be difficult, in particular regarding the violation of the independence from irrelevant alternatives (IIA) assumption. Finally, in this setting, several countries can be treated several times and at different periods. To the best of my knowledge there doesn’t exist a difference-in-differences

[24] See table 2.H.1 for more information on each category.

	All mig. pol.	Restrictive mig. pol.	Permissive mig. pol.
All pop. leaders			
Populism	0.097	0.332	-0.044
	(0.202)	(0.313)	(0.245)
RW pop. leaders			
Populism	0.126	0.582*	-0.200
	(0.239)	(0.300)	(0.302)
LW pop. leaders			
Populism	-0.012	-0.651	0.217
	(0.389)	(0.843)	(0.430)
Year FE	Yes	Yes	Yes
Country FE	Yes	Yes	Yes
Obs.	790	660	758

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Heteroskedasticity-robust standard errors in parenthesis. Only major policy changes (according to DEMIG classification) are kept. The dependent variable is the number of migration policies, the number of more restrictive migration policies only, and the number of more permissive migration policies only. The interest variable is an indicator of populism, considering all populist leaders in the first row, right-wing populist leaders only in the second row, and left-wing populist leaders only in the third row.

Table 2.5.2: Effect of populism on the number of major migration policies (according to DEMIG's classification).

variant that is able to take all those features into account.

Therefore, I only provide suggestive evidence on the potential effect of populism on the attractiveness of a country. I compute the number of respondents who would like to migrate to each country every year, and I rank those destinations based on this count (having a lower rank means higher attractiveness). I focus on all countries that experienced a populist leader at some point between 2008 and 2018. Figure 2.5.1 shows the evolution of the rank for those countries. At first glance, it is difficult to see a common pattern when a populist arrives at power: in some cases (e.g South Africa in 2009) there is a sharp increase in attractiveness, but in others (e.g Philippines 2016) we see a decline. It is also worth noting that several countries (e.g USA, Italy, India) have a remarkably stable rank, no matter who is at power.

As said above, this graph doesn't tell us anything about causality. Still, it shows that there is no obvious impact of populism on the attractiveness of a country. One might explain this by the way Gallup questions are phrased. Indeed, the question I used specifically mention "migrate permanently". Hence, it is possible that respondents do care about the arrival of a populist at power in the country they want to move to, but consider that this won't last too much time and that they could migrate afterwards.

Figure 2.5.1: Change in attractiveness of countries with a populist leader. The y-axis shows the rank of each destination per year (a lower value means a higher attractiveness). Countries are ranked based on the number of respondents who would like to migrate permanently to this country. The three lines correspond to the rank when we use the answers of all respondents, high-skilled respondents only, and low-skilled respondents only. The shaded area shows the periods during which a left-/right-wing populist leader was at power.

2.5.3 Returning to the home country

In the previous section, I focused on people who have a desire to migrate, but didn't move yet. In this section, I explore whether people who already migrated are affected by having a populist leader in their destination country. This group could feel directly threatened by this political change and hence decide to voluntarily go back to their home country.

To analyze whether this is the fact, I keep using the Gallup World Polls but focus only respondents who are born in another country than the one in which they currently live. I estimate the following logit model:

$$Pr(Y_{ioct} = 1|Pop_{ct}, X_i) = \alpha + \beta Pop_{ct} + \gamma \mathbf{X_i} + \omega Pop_{ot} + \nu_c + \mu_o + \tau_t + \varepsilon_{ioct} \tag{2.4}$$

where Y_{ioct} is a dummy variable equal to 1 if individual i that is born abroad in country o and living in country c at time t expresses a desire to go back to their origin country. Pop_{ct} is a dummy indicating whether there is a populist leader at power in country c at year t, and $\mathbf{X_i}$ is a vector of control variables at the individual level (age, gender, and income in quintiles). I also include a dummy indicating whether there is a populist leader at origin (Pop_{ot}). Finally, I include fixed effects for the country in which migrants live (ν_c), their origin country (μ_o), and time fixed effects (τ_t).

Results in tables 2.I.1-2.I.4 in the Appendix show the results. If we look at migrants in general (table 2.I.1), then having a populist leader doesn't seem to have any effect on the willingness to go back to the home country. However, the story is very different when we distinguish respondents by their skill level (tables 2.I.2-2.I.4). Indeed, high-skilled and middle-skilled migrants are mostly indifferent to having a populist leader, regardless of the political side of this leader. On the other hand, while low-skilled migrants don't show a higher desire to either stay or go back home when a left-wing populist is at power, their willingness to leave the country increases when a right-wing populist is the head of the country. This is consistent with the results from section 2.5.1 showing that right-wing populist leaders implement more restrictive migration policies. Those policies cover migrants' integration and access to the labor market, and those harder conditions probably play an important role when one arbitrates whether they should stay in the country or go back in their origin country.

2.6 Conclusion

The recent increase in populism is accompanied by a large literature on its causes, with an important part of it focusing on immigration. However, the potential effects of populism on immigration are largely understudied in comparison, due to the difficulty of finding quasi-natural experiments or suitable counterfactuals to countries having a populist leader.

In this paper, I estimate the effect of having a populist leader on the growth rate of immigration using data on the last 60 years. Using the generalized synthetic control, I find that having a populist leader has a strong negative effect on the growth rate of immigration, and in particular on low-skilled immigration. After three years, the immigration growth rate is almost 20 percentage points lower than in the absence of a populist leader. However, emigration is almost left unchanged by the populist leader.

Turning to the mechanisms behind those results, I use data from the DEMIG policy project on a large number of migration policies to explore the political consequences of populism. I find that right-wing populist leaders implement more restrictive immigration policies than left-wing leaders, targeting specifically low-skilled migrants. I also explore the relationship between populism and migration intentions using data from the Gallup World Polls. I don't find any clear evidence that having a populist leader deteriorates the country's attractiveness.

However, I show that low-skilled migrants have more intentions to leave the country to go back home when a right-wing populist leader is at power, while middle-skilled migrants and high-skilled migrants are less affected by this. Therefore, the negative effect on immigration almost exclusively comes from stricter immigration policies.

Finally, my results show that the effect of populism on international migration are quite heterogeneous across countries and years. Hence, it is important to continue studying those events on a case-by-case basis.

Appendix

2.A Populist events

Country	Year	Leader	Left/Right
Argentina	1973	Perón-Martínez	Left
Argentina	1989	Menem	Right
Argentina	2003	Kirchner-Fernández	Left
Brazil	1990	Collor	Right
Ecuador	1996	Bucaram	Right
Greece	1981	Papandreou	Left
Greece	1993	Papandreou	Left
Japan	2001	Koizumi	Right
Philippines	1998	Estrada	Left
South Korea	2003	Roh	Right
Thailand	2001	Shinawatra	Right
Turkey	2003	Erdoğan	Right

Table 2.A.1: List of populist events used in the synthetic control method.

2.B Variants of the synthetic control method

I use various methods to combine the multiple synthetic controls (one per event). Note that to the best of my knowledge, those methods do not allow to derive a confidence interval (contrarily to the generalized synthetic control). Therefore, their objective is mostly to check that the direction and the magnitude of the aggregate effect found with the generalized synthetic control is consistent with alternative methods.

Aggregate separate SCM. First, I apply the SCM to each event and then average the results. I use the fit in the pre-treatment period to compute a weight for each event. The idea is to give more weight to events where the fit in the pre-treatment period is good, i.e. the root mean squared prediction error (RMSPE) in this period is low. I compute the weight of each event σ_i as:

$$RMSPE_i = \frac{\sum_{t \in T} [Y_{it} - \hat{Y_{it}}]^2}{T}$$

$$\sigma_i = \frac{1}{RMSPE_i}$$

where Y_{it} and $\hat{Y_{it}}$ are respectively the outputs of the treated country and of the synthetic control in event i at time t, and T is the length of the pre-treatment period. Therefore, the higher the difference between the two trends in the pre-treatment period (i.e. the worse the fit), the lower the weight attributed to the event. I then take the weighted mean of these events at each period.

Partially pooled SCM. Aggregating synthetic controls applied on separate events means that I only care about imbalance at the event level, not at the aggregate level. Another way to estimate the average effect of populism on immigration is to use pooled SCM by gathering all events in a single large panel dataset centered around t_0, and then by applying a single SCM on this pooled dataset. While this method requires only one SCM, it completely ignores the imbalances at the event level. One middle ground between these two methods is to use the partially pooled SCM (Ben-Michael et al., 2021). In this method, the SCM weights are chosen so that they minimize a weighted average of the pooled and unit-specific pre-treatment fits. Ben-Michael et al. (2021) note ν and $1 - \nu$ the weight given to reducing the imbalance of the pooled SCM and of the separate SCM respectively. When $\nu = 0$, I are in the case of the separate SCM as above and when $\nu = 1$, I are in the case of the pooled SCM (not used here). Therefore, the idea of partially pooled SCM is to find the value of ν that minimizes this weighted average of separate and pooled imbalances. Figure 2.B.1 how both types of imbalances change with the value of ν. I can see that moving from $\nu = 0$ (as is the case when I do the separate SCM) to $\nu = 0.95$ allows me to greatly reduce the pooled imbalance while slightly increasing the event-level imbalance. Therefore, I use this value of ν when I apply the partially pooled SCM. I use the R package `augsynth` (v.0.2.0, Ben-Michael, 2022).

Synthetic difference-in-differences. Finally, I also apply the synthetic difference-in-differences (SDiD, Arkhangelsky et al., 2021). This method is a mixture of the synthetic control method and of the difference-in-differences method. It still computes unit weights to create a synthetic control, but it is no longer necessary for it to fit very closely the treated unit in the pre-treatment period. Instead, it requires the trends of the treated unit and of the synthetic control to be parallel in the pre-treatment period. Moreover, this method also computes time weights to balance the pre-treatment and post-treatment outcomes for the control units. Specifically, ”time weights are designed so that the average posttreatment outcome for each of the control units differs by a constant from the weighted average of the pretreatment outcomes for the same control units” (p.4090, Arkhangelsky et al., 2021). I used the R package `synthdid` (v.0.0.9, Arkhangelsky, 2023).

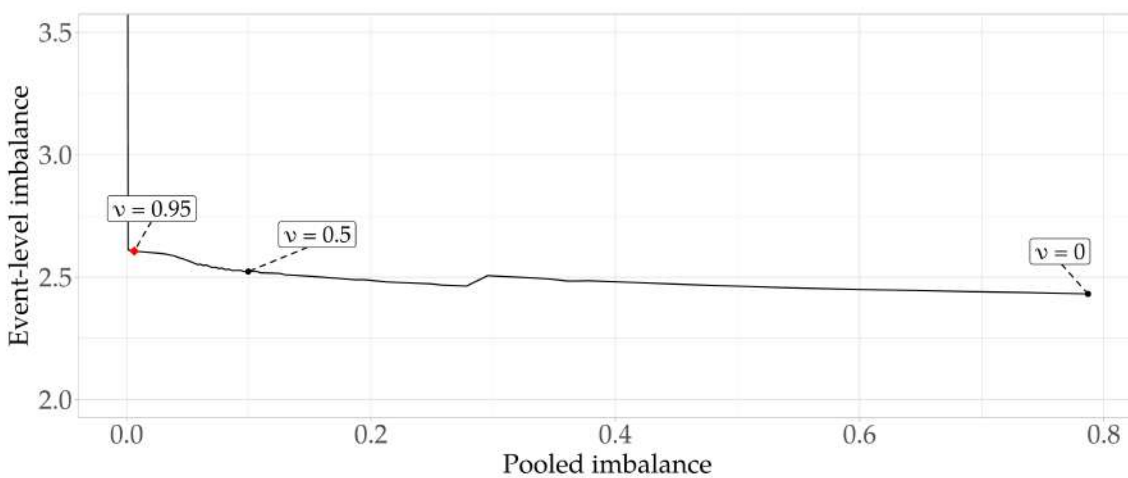


Figure 2.B.1: Balance possiblity frontier. This shows how the event-level and the pooled imbalance change with the value of ν in the partially pooled SCM.

2.C SCM with time placebo

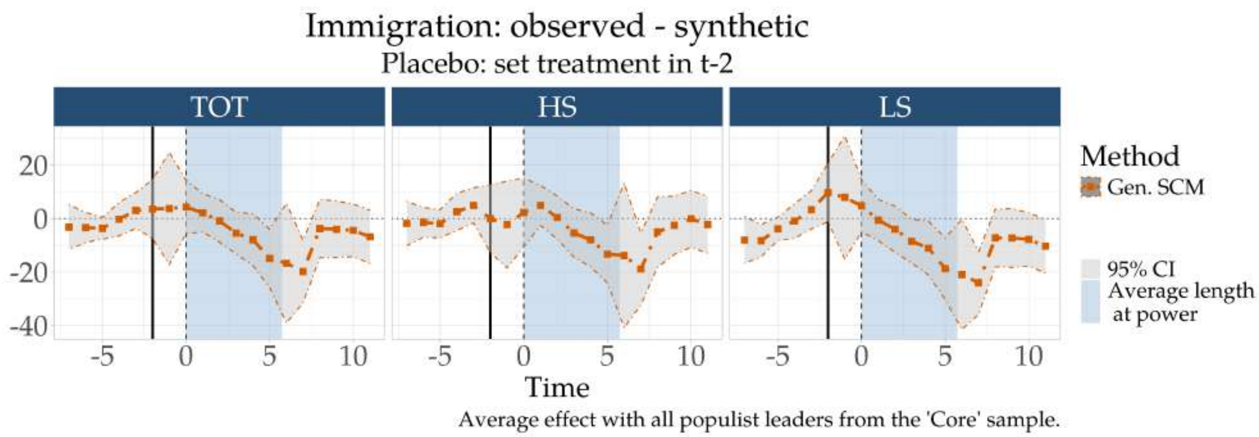


Figure 2.C.1: Effect of having a populist leader on growth rate of immigration (in percentage points). The orange dashed line shows the difference between the observed growth rate of immigration and the synthetic one. The vertical solid black line shows the fake treatment time. The gray band shows the 95% CI, computed with bootstrap (1,000 repetitions).

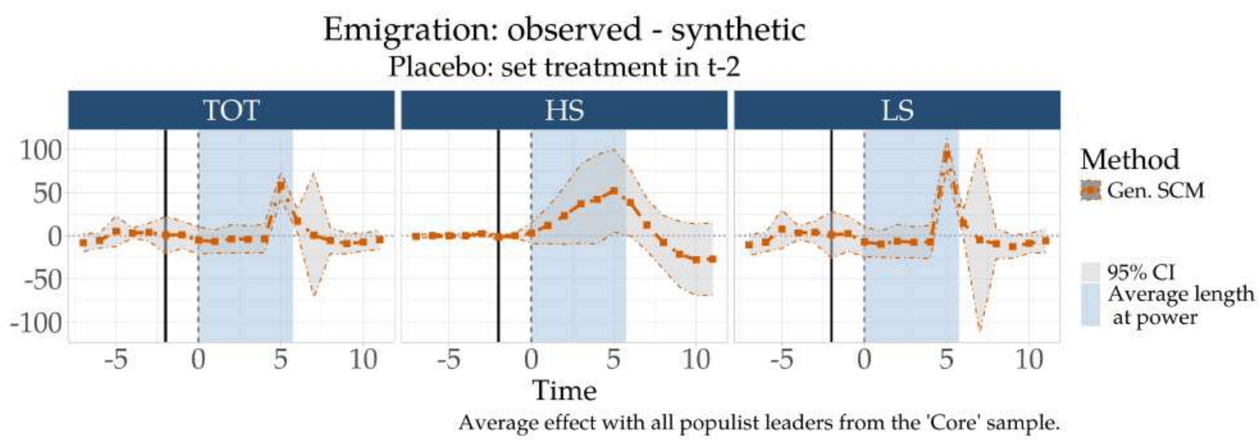


Figure 2.C.2: Effect of having a populist leader on growth rate of emigration (in percentage points). The orange dashed line shows the difference between the observed growth rate of immigration and the synthetic one. The vertical solid black line shows the fake treatment time. The gray band shows the 95% CI, computed with bootstrap (1,000 repetitions).

2.D SCM focusing on scholars migration only

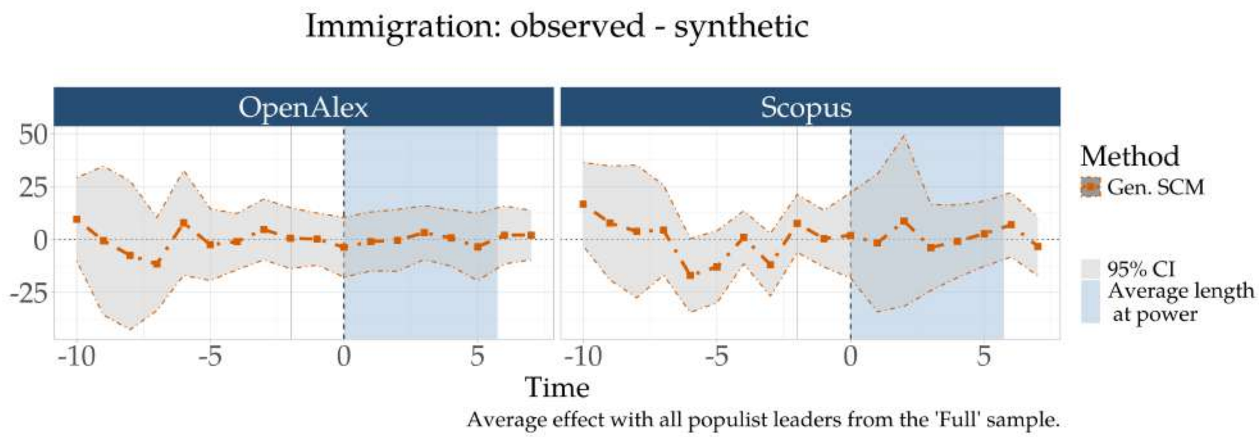


Figure 2.D.1: Effect of having a populist leader on growth rate of immigration of scholars (in percentage points). The orange dashed line shows the difference between the observed growth rate of immigration and the synthetic one. The gray band shows the 95% CI, computed with bootstrap (1,000 repetitions).

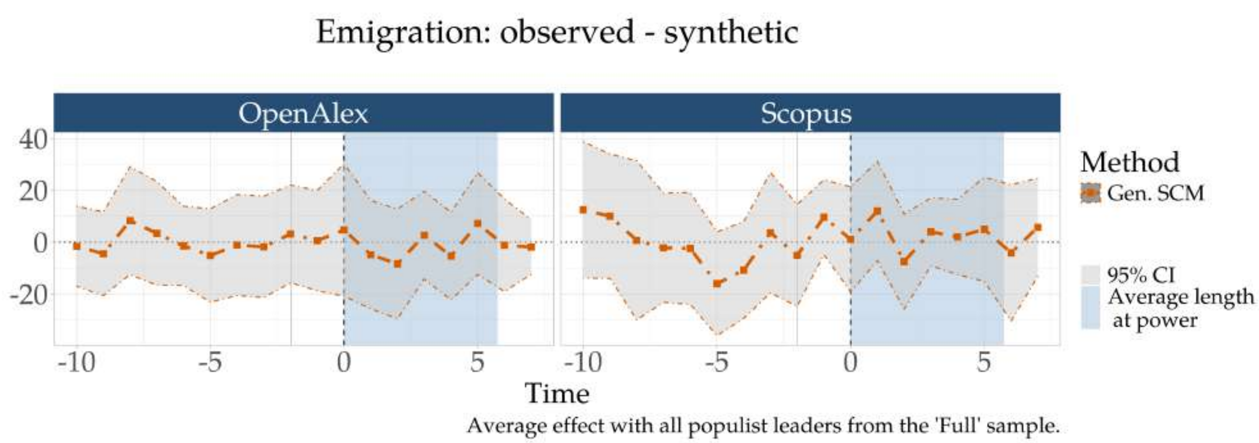


Figure 2.D.2: Effect of having a populist leader on growth rate of emigration of scholars (in percentage points). The orange dashed line shows the difference between the observed growth rate of immigration and the synthetic one. The gray band shows the 95% CI, computed with bootstrap (1,000 repetitions).

2.E SCM without Ecuador 1996

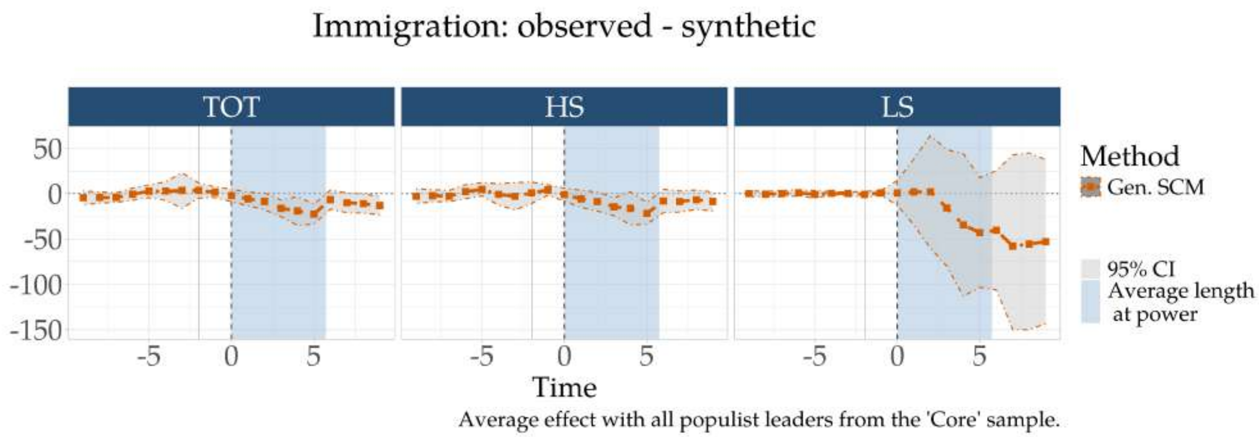


Figure 2.E.1: Effect of having a populist leader on growth rate of immigration (in percentage points). The orange dashed line shows the difference between the observed growth rate of immigration and the synthetic one. The gray band shows the 95% CI, computed with bootstrap (1,000 repetitions). This result contains all events except Ecuador 1996.

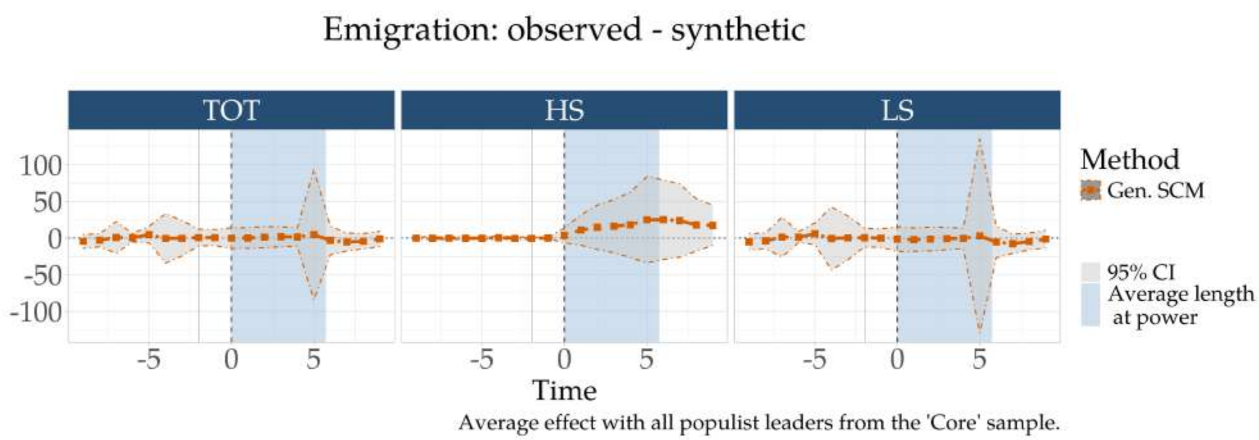


Figure 2.E.2: Effect of having a populist leader on growth rate of emigration (in percentage points). The orange dashed line shows the difference between the observed growth rate of immigration and the synthetic one. The gray band shows the 95% CI, computed with bootstrap (1,000 repetitions). This result contains all events except Ecuador 1996.

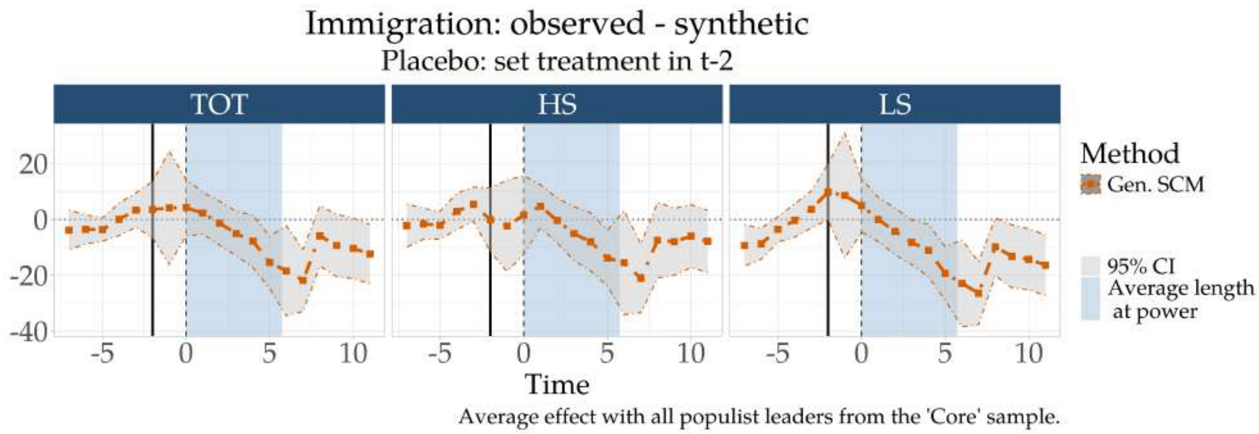


Figure 2.E.3: Effect of having a populist leader on growth rate of immigration (in percentage points). The orange dashed line shows the difference between the observed growth rate of immigration and the synthetic one. The vertical solid black line shows the fake treatment time. The gray band shows the 95% CI, computed with bootstrap (1,000 repetitions). This result contains all events except Ecuador 1996.

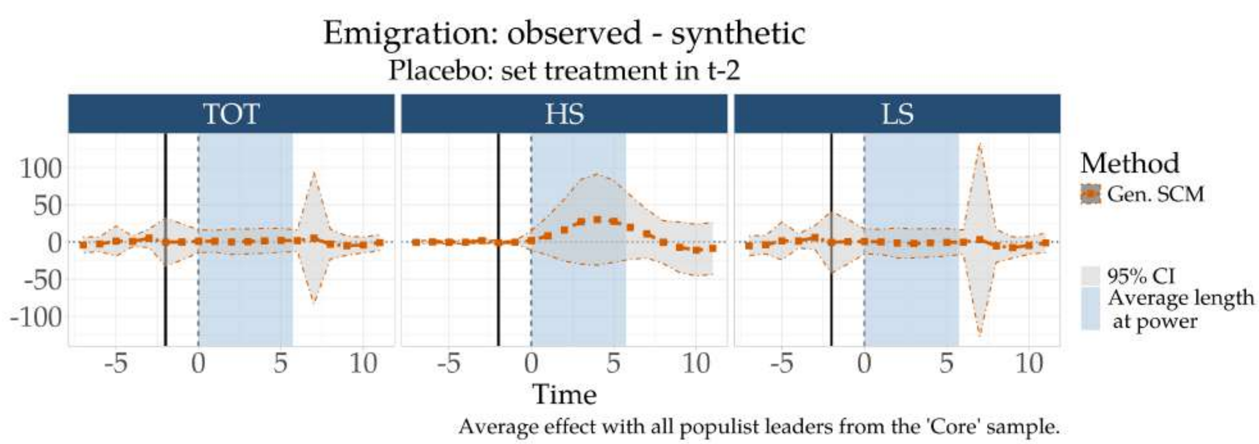


Figure 2.E.4: Effect of having a populist leader on growth rate of emigration (in percentage points). The orange dashed line shows the difference between the observed growth rate of immigration and the synthetic one. The vertical solid black line shows the fake treatment time. The gray band shows the 95% CI, computed with bootstrap (1,000 repetitions). This result contains all events except Ecuador 1996.

2.F Results with variants of SCM

2.F.1 Immigration

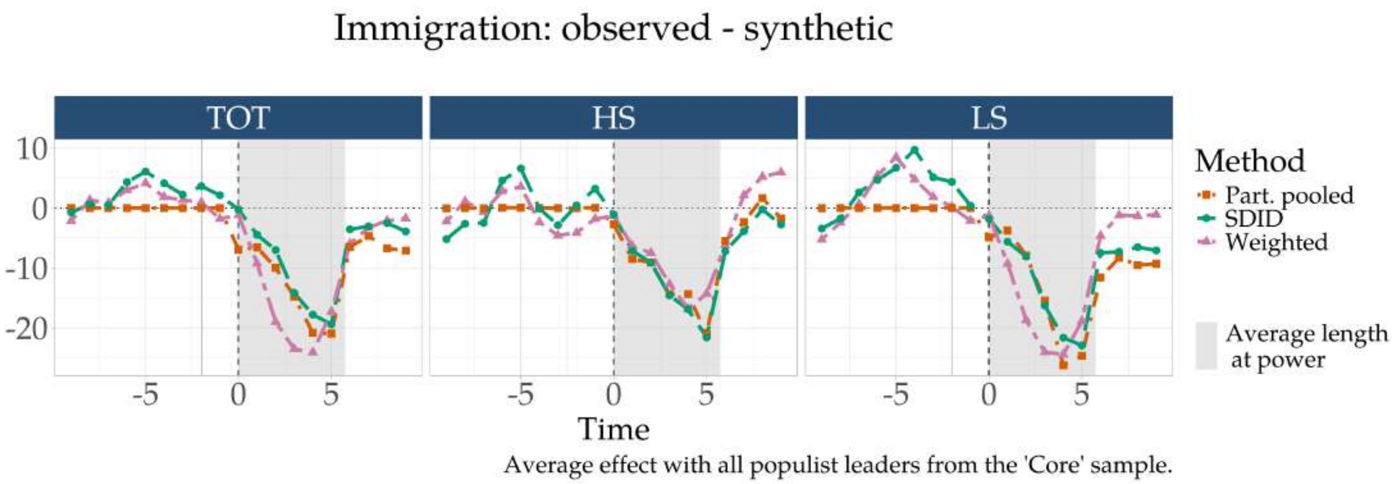


Figure 2.F.1: This figure has the same setting as figure 2.4.1 but uses the alternative methods described in appendix 2.B.

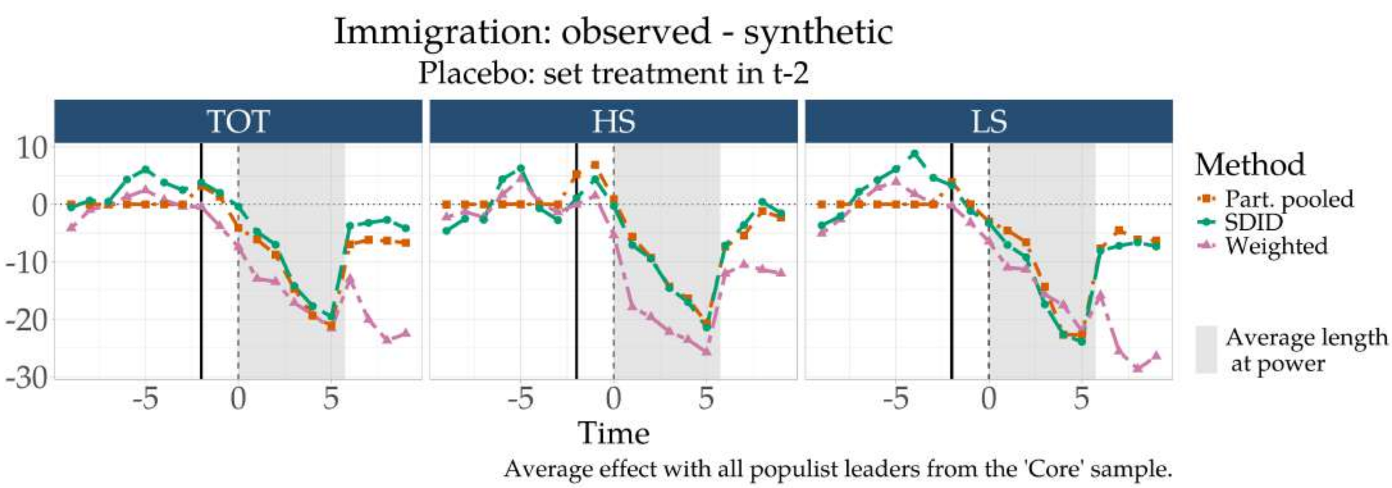


Figure 2.F.2: This figure has the same setting as figure 2.F.1 but has treatment in t_{-2}.

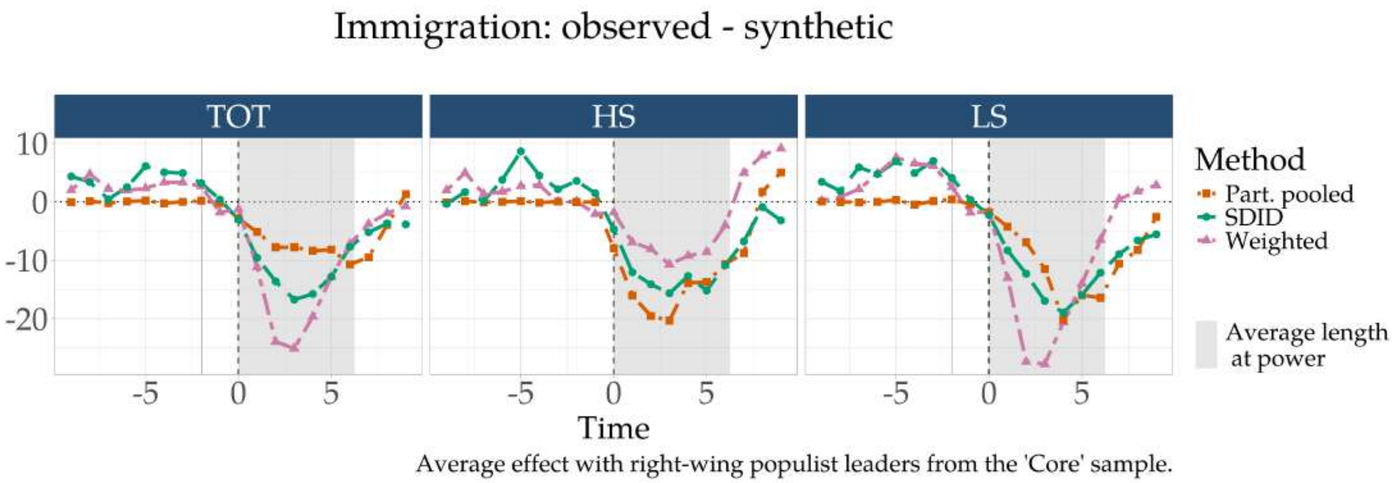


Figure 2.F.3: This figure has the same setting as figure 2.4.5 but uses the alternative methods described in appendix 2.B.

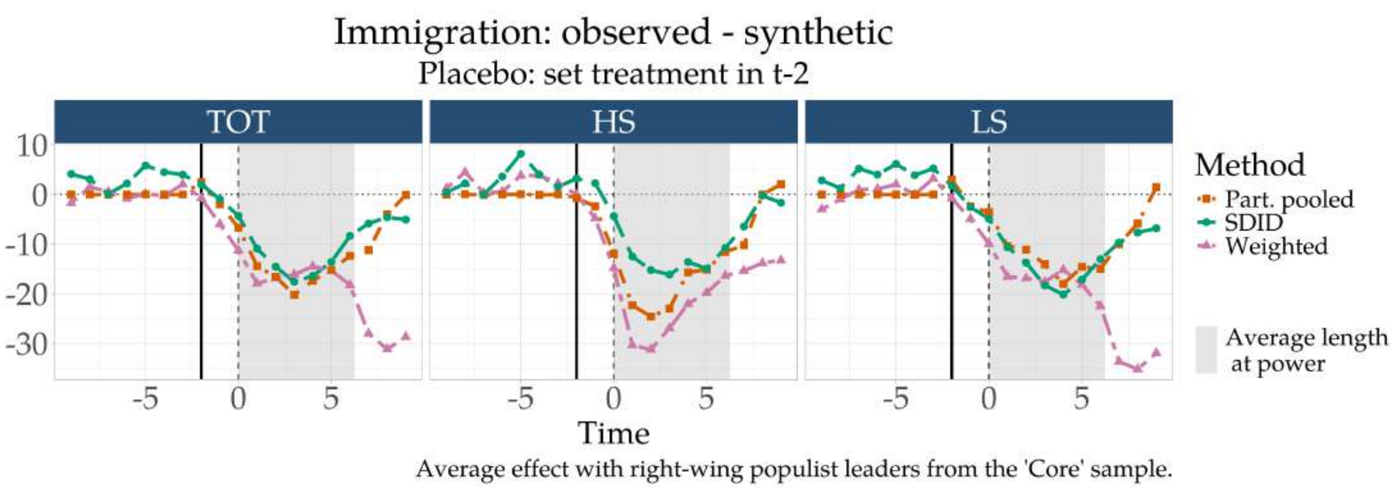


Figure 2.F.4: This figure has the same setting as figure 2.F.3 but has treatment in t_{-2}.

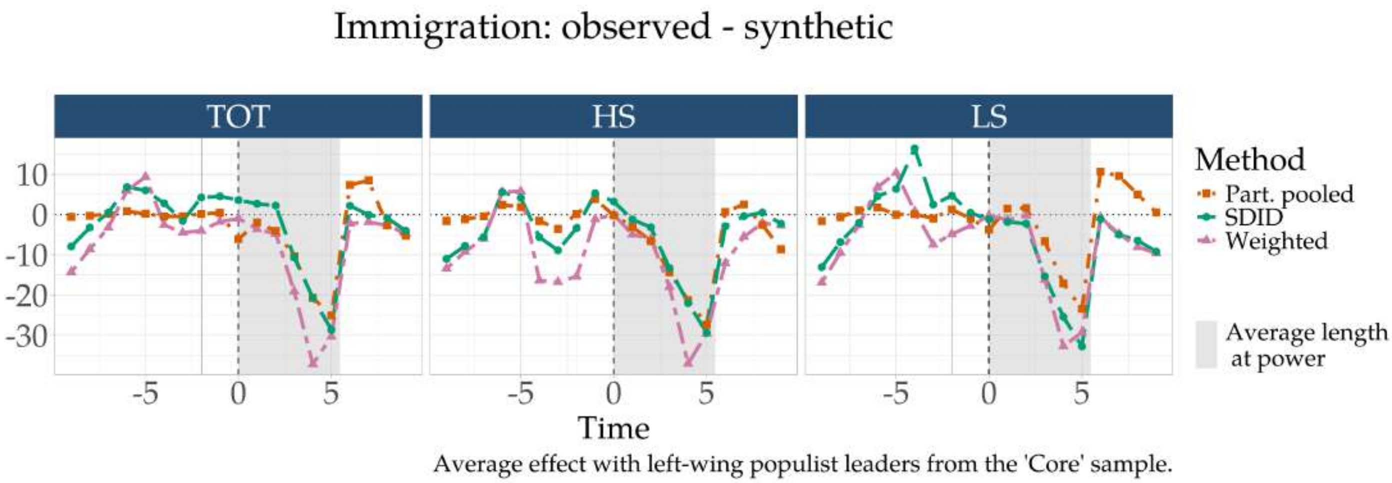


Figure 2.F.5: This figure has the same setting as figure 2.4.6 but uses the alternative methods described in appendix 2.B.

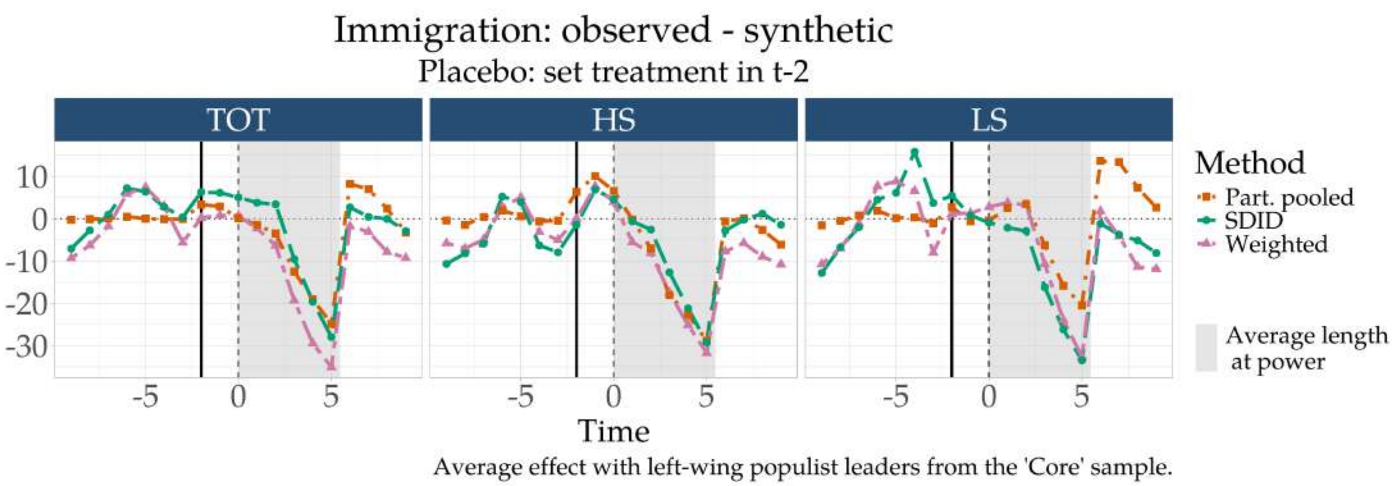


Figure 2.F.6: This figure has the same setting as figure 2.F.5 but has treatment in t_{-2}.

2.F.2 Emigration

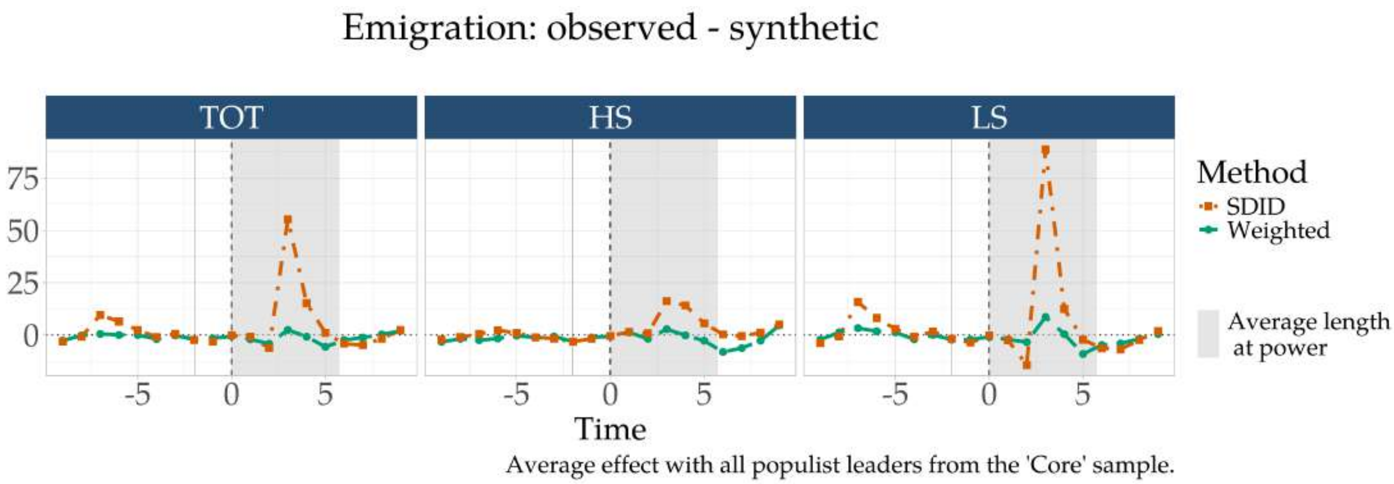


Figure 2.F.7: This figure has the same setting as figure 2.4.3 but uses the alternative methods described in appendix 2.B.

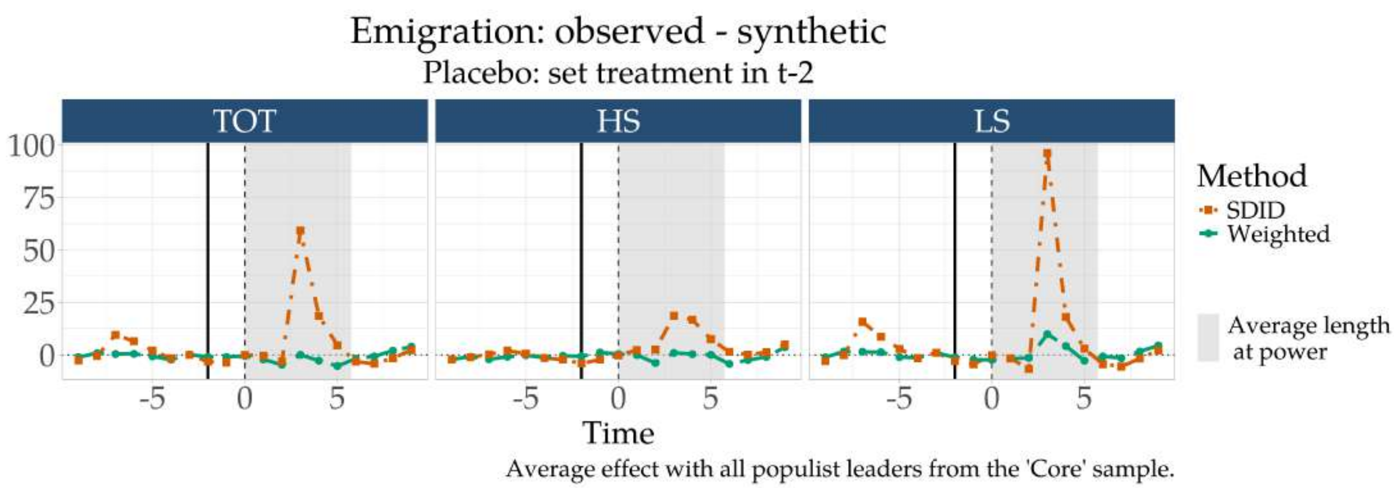


Figure 2.F.8: This figure has the same setting as figure 2.F.7 but has treatment in t_{-2}.

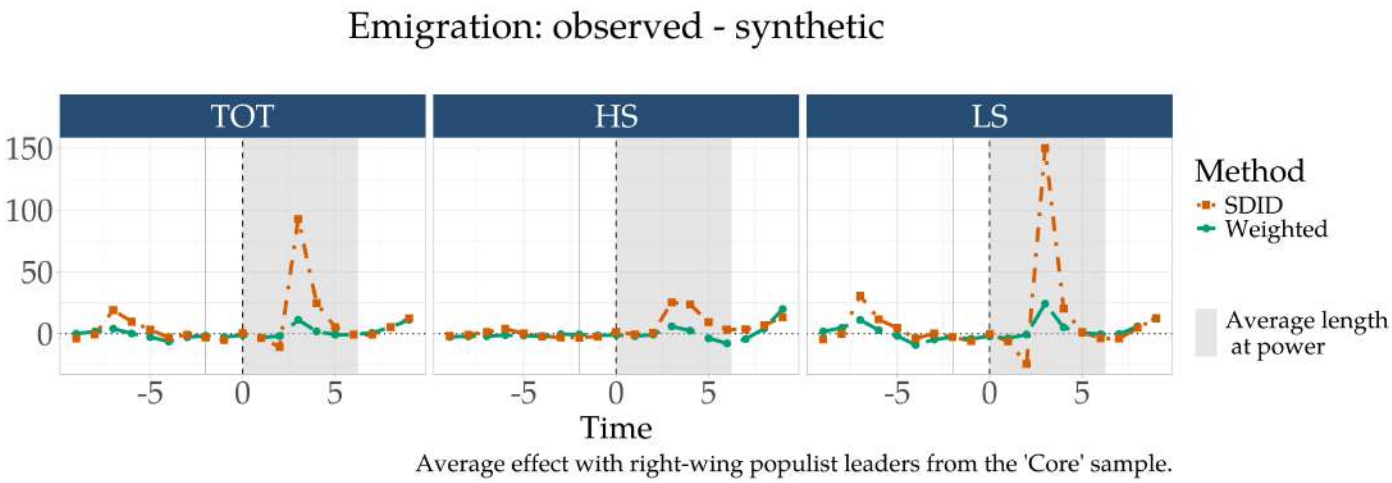


Figure 2.F.9: This figure has the same setting as figure 2.4.5 but uses the alternative methods described in appendix 2.B.

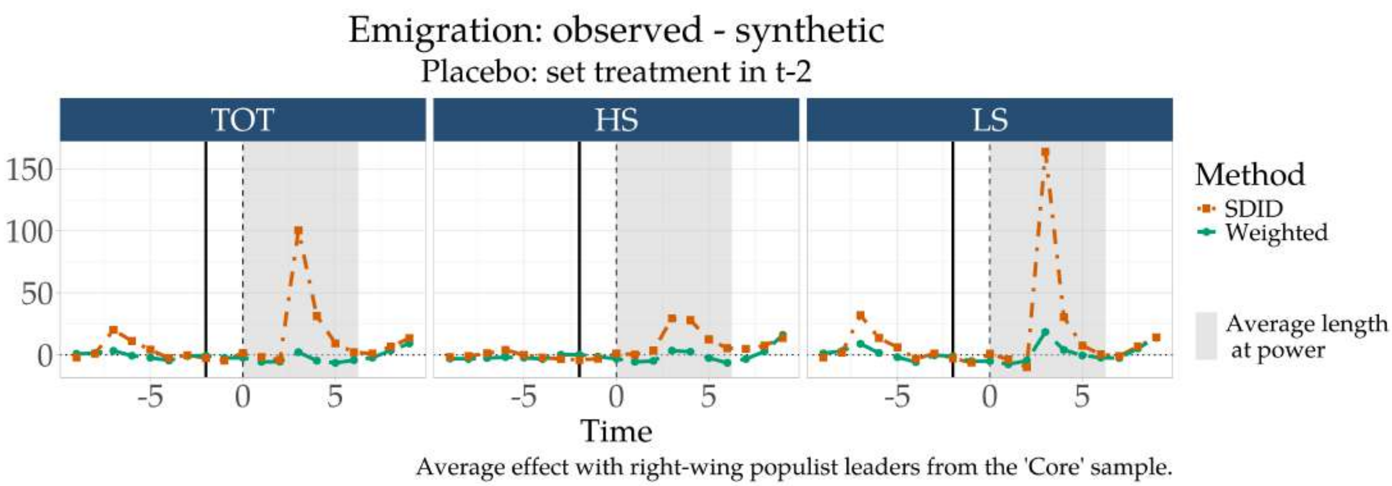


Figure 2.F.10: This figure has the same setting as figure 2.F.9 but has treatment in t_{-2}.

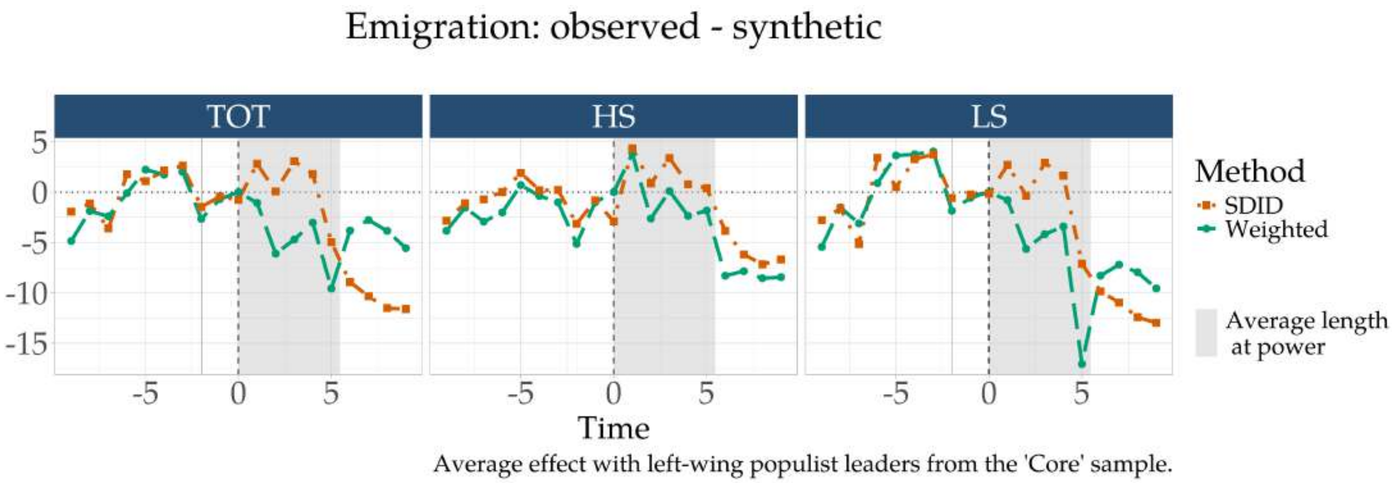


Figure 2.F.11: This figure has the same setting as figure 2.4.6 but uses the alternative methods described in appendix 2.B.

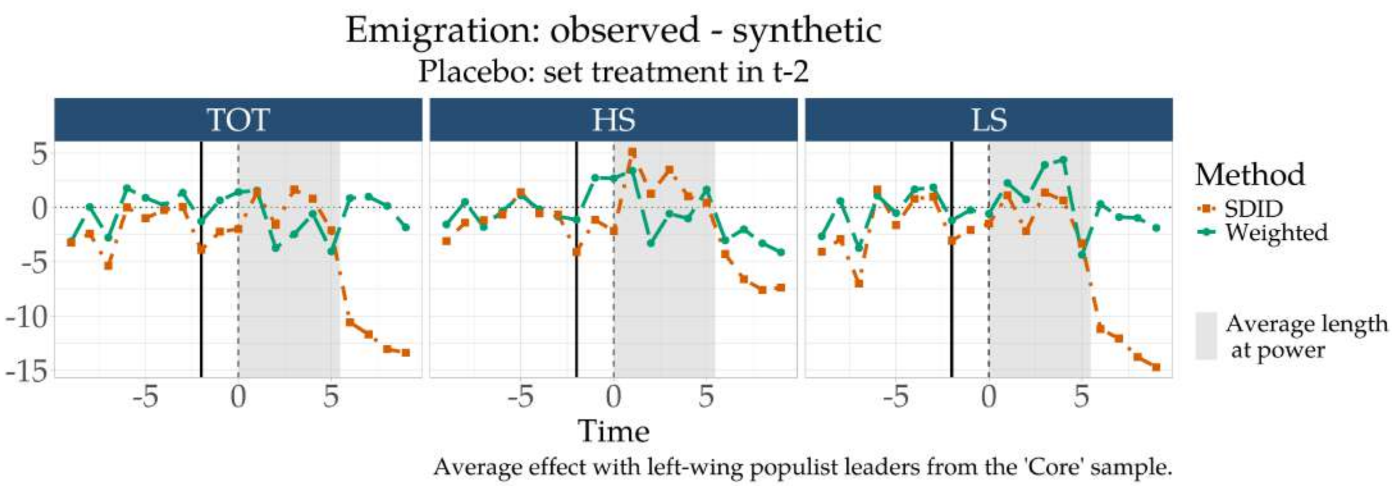


Figure 2.F.12: This figure has the same setting as figure 2.F.11 but has treatment in t_{-2}.

2.G Additional results for the migration policy analysis

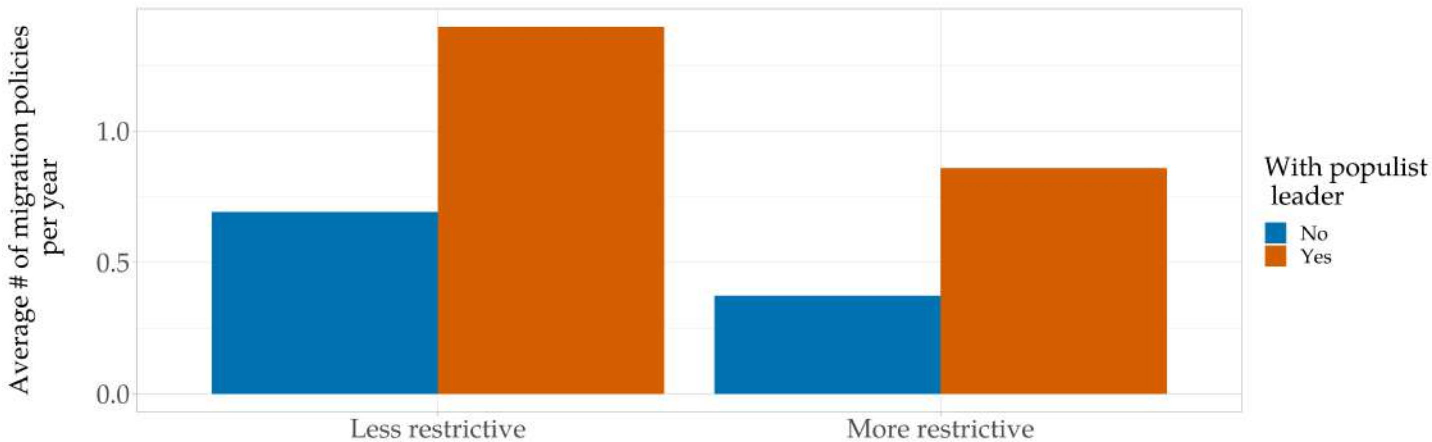


Figure 2.G.1: Average number of migration policy changes per year and per effect on restrictiveness.

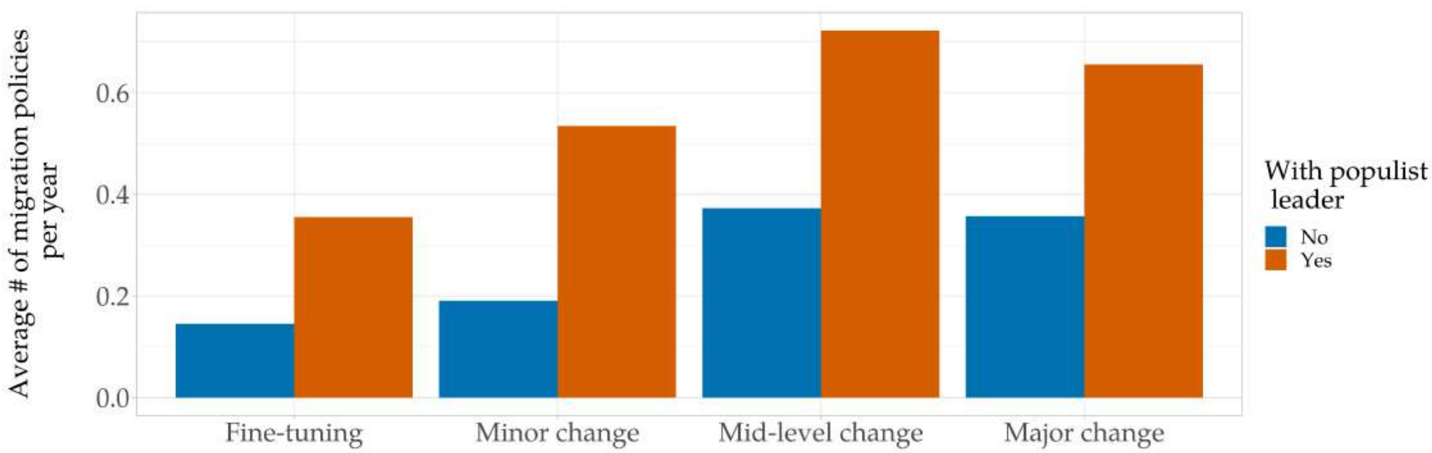


Figure 2.G.2: Average number of migration policy changes per year and per importance of the change.

2.H DEMIG policy information

Fine-tuning change	Fine-tuning changes are measures that only affect part of a migrant category and only alter an existing policy instrument.
Minor change	Minor changes are measures that affect an entire migrant category, and only alter an existing policy instrument.
Mid-level change	Mid-level changes are measures that only affect part of a migrant category, but introduce or remove a new policy instrument.
Major change	Major changes are measures that affect an entire migrant category and introduce or remove a new policy instrument.

Table 2.H.1: Possible values of the "magnitude of change" variable in the DEMIG policy dataset. This comes from the DEMIG policy codebook.

2.I On return migration flows

	Want to go back home		
	(1)	(2)	(3)
Populism	-0.1367		
	(0.1763)		
L-W Populism		-0.1179	
		(0.2117)	
R-W Populism			-0.1888
			(0.3311)
Age	-0.0167***	-0.0167***	-0.0167***
	(0.0017)	(0.0017)	(0.0017)
Income	-0.0672***	-0.0672***	-0.0672***
	(0.0187)	(0.0187)	(0.0186)
Gender	0.0429	0.0430	0.0429
	(0.0540)	(0.0540)	(0.0540)
Populism at origin	-0.0599	-0.0610	-0.0625
	(0.1467)	(0.1468)	(0.1466)
Fixed-effects			
Origin country	Yes	Yes	Yes
Country	Yes	Yes	Yes
Year	Yes	Yes	Yes
Fit statistics			
Observations	22,578	22,578	22,578
Log-Likelihood	-5,437.5	-5,437.6	-5,437.6
Pseudo R^2	0.106	0.106	0.106

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*
The sample of interest consists of Gallup respondents who were born in another country. The dependent variable is a dummy equal to one if the respondent wants to leave the country and go back home, and 0 otherwise. The interest variables are dummies indicating whether there is a populist at power (any populist, left-wing only, right-wing only). Other control variables are the gender (1 if female, 0 otherwise), income, and age of the individual. This table includes fixed effects for the country in which migrants live, for the country they come from, and for the year.

Table 2.I.1: Effect of having a populist leader on the likelihood to return to the home country (all migrants).

	Want to go back home		
	(1)	(2)	(3)
Populism	-0.6737		
	(0.4907)		
L-W Populism		-0.6422	
		(0.6231)	
R-W Populism			-0.7732
			(0.8430)
Age	-0.0176***	-0.0176***	-0.0177***
	(0.0035)	(0.0035)	(0.0035)
Income	-0.0353	-0.0351	-0.0352
	(0.0356)	(0.0356)	(0.0356)
Gender	0.2664***	0.2680***	0.2631***
	(0.0981)	(0.0982)	(0.0978)
Populism at origin	0.2046	0.2025	0.2061
	(0.2575)	(0.2587)	(0.2560)
Fixed-effects			
Origin country	Yes	Yes	Yes
Country	Yes	Yes	Yes
Year	Yes	Yes	Yes
Fit statistics			
Observations	7,782	7,782	7,782
Log-Likelihood	-1,712.0	-1,712.5	-1,712.7
Pseudo R^2	0.117	0.116	0.116

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*
The sample of interest consists of high-skilled Gallup respondents who were born in another country. The dependent variable is a dummy equal to one if the respondent wants to leave the country and go back home, and 0 otherwise. The interest variables are dummies indicating whether there is a populist at power (any populist, left-wing only, right-wing only). Other control variables are the gender (1 if female, 0 otherwise), income, and age of the individual. This table includes fixed effects for the country in which migrants live, for the country they come from, and for the year.

Table 2.I.2: Effect of having a populist leader on the likelihood to return to the home country (high-skilled migrants).

	Want to go back home		
	(1)	(2)	(3)
Populism	-0.2232		
	(0.2159)		
L-W Populism		-0.1646	
		(0.2569)	
R-W Populism			-0.4128
			(0.4263)
Age	-0.0148***	-0.0148***	-0.0148***
	(0.0023)	(0.0023)	(0.0023)
Income	-0.0773***	-0.0772***	-0.0776***
	(0.0257)	(0.0257)	(0.0257)
Gender	-0.0823	-0.0820	-0.0818
	(0.0729)	(0.0729)	(0.0729)
Populism at origin	-0.2718	-0.2746	-0.2778
	(0.2013)	(0.2013)	(0.2010)
Fixed-effects			
Origin country	Yes	Yes	Yes
Country	Yes	Yes	Yes
Year	Yes	Yes	Yes
Fit statistics			
Observations	11,896	11,896	11,896
Log-Likelihood	-2,957.4	-2,957.7	-2,957.4
Pseudo R^2	0.118	0.118	0.118

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*
The sample of interest consists of middle-skilled Gallup respondents who were born in another country. The dependent variable is a dummy equal to one if the respondent wants to leave the country and go back home, and 0 otherwise. The interest variables are dummies indicating whether there is a populist at power (any populist, left-wing only, right-wing only). Other control variables are the gender (1 if female, 0 otherwise), income, and age of the individual. This table includes fixed effects for the country in which migrants live, for the country they come from, and for the year.

Table 2.I.3: Effect of having a populist leader on the likelihood to return to the home country (middle-skilled migrants).

	Want to go back home		
	(1)	(2)	(3)
Populism	1.099**		
	(0.5440)		
L-W Populism		0.7484	
		(0.5705)	
R-W Populism			1.702*
			(0.9361)
Age	-0.0204***	-0.0202***	-0.0204***
	(0.0050)	(0.0050)	(0.0050)
Income	0.0163	0.0167	0.0097
	(0.0620)	(0.0621)	(0.0622)
Gender	0.0454	0.0455	0.0482
	(0.1671)	(0.1666)	(0.1668)
Populism at origin	0.5655	0.5700	0.5863
	(0.5009)	(0.5069)	(0.5050)
Fixed-effects			
Origin country	Yes	Yes	Yes
Country	Yes	Yes	Yes
Year	Yes	Yes	Yes
Fit statistics			
Observations	2,202	2,202	2,202
Log-Likelihood	-627.5	-629.1	-627.9
Pseudo R^2	0.128	0.126	0.128

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*
The sample of interest consists of low-skilled Gallup respondents who were born in another country. The dependent variable is a dummy equal to one if the respondent wants to leave the country and go back home, and 0 otherwise. The interest variables are dummies indicating whether there is a populist at power (any populist, left-wing only, right-wing only). Other control variables are the gender (1 if female, 0 otherwise), income, and age of the individual. This table includes fixed effects for the country in which migrants live, for the country they come from, and for the year.

Table 2.I.4: Effect of having a populist leader on the likelihood to return to the home country (low-skilled migrants).

References

Abadie, A. (2021). "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects". In: *Journal of Economic Literature* 59(2), pp. 391–425.

Akbaritabar, A., T. Theile, et al. (2024). "Bilateral Flows and Rates of International Migration of Scholars for 210 Countries for the Period 1998-2020". In: *Scientific Data* 11(1), p. 816.

Arkhangelsky, D. (2023). *Synthdid: Synthetic Difference-in-Difference Estimation*. manual.

Arkhangelsky, D., S. Athey, et al. (2021). "Synthetic Difference-in-Differences". In: *American Economic Review* 111(12), pp. 4088–4118.

Arslan, C., J.-C. Dumont, et al. (2015). *A New Profile of Migrants in the Aftermath of the Recent Economic Crisis*. OECD Social, Employment and Migration Working Papers 160.

Artuc, E., F. Docquier, et al. (2015). "A Global Assessment of Human Capital Mobility: The Role of Non-OECD Destinations". In: *World Development* 65, pp. 6–26.

Barone, G., A. D'Ignazio, et al. (2016). "Mr. Rossi, Mr. Hu and Politics. The Role of Immigration in Shaping Natives' Voting Behavior". In: *Journal of Public Economics* 136, pp. 1–13.

Bellodi, L., F. Docquier, et al. (2024). *Digging up Trenches: Populism, Selective Mobility, and the Political Polarization of Italian Municipalities*. IZA Discussion Paper Series 16732.

Ben-Michael, E. (2022). *Augsynth: The Augmented Synthetic Control Method*.

Ben-Michael, E., A. Feller, et al. (2021). "The Augmented Synthetic Control Method". In: *Journal of the American Statistical Association* 116(536), pp. 1789–1803.

Bracco, E., M. De Paola, et al. (2018). "The Effect of Far Right Parties on the Location Choice of Immigrants: Evidence from Lega Nord Mayors". In: *Journal of Public Economics* 166, pp. 12–26.

Cerqua, A. and F. Zampollo (2023). "Deeds or Words? The Local Influence of Anti-Immigrant Parties on Foreigners' Flows". In: *European Journal of Political Economy* 77, p. 102275.

Chen, S. (2023). *Unemployment, Immigration, and Populism*. IZA Discussion Paper Series 16642.

DEMIG (2015). *DEMIG POLICY*. Online Edition. Version 1.3.

Demirci, M. (2023). "Youth Responses to Political Populism: Education Abroad as a Step toward Emigration". In: *Journal of Comparative Economics* 51(2), pp. 653–673.

Docquier, F., L. Guichard, et al. (2023). *Populism and the Skill-Content of Globalization*.

Docquier, F. and C. Vasilakis (2023). *Migrants' Self-Selection and the Vicious Circle of Right-Wing Populism*.

Edo, A., Y. Giesing, et al. (2019). "Immigration and Electoral Support for the Far-Left and the Far-Right". In: *European Economic Review* 115, pp. 99–143.

Facchini, G. and A. M. Mayda (2012). "Individual Attitudes Towards Skilled Migration: An Empirical Analysis Across Countries". In: *The World Economy* 35(2), pp. 183–196.

Funke, M., M. Schularick, et al. (2023). "Populist Leaders and the Economy". In: *American Economic Review* 113(12), pp. 3249–3288.

Guriev, S. and E. Papaioannou (2022). “The Political Economy of Populism”. In: *Journal of Economic Literature* 60(3), pp. 753–832.

Hainmueller, J. and M. J. Hiscox (2010). “Attitudes toward Highly Skilled and Low-skilled Immigration: Evidence from a Survey Experiment”. In: *The American Political Science Review* 104(1), pp. 61–84. JSTOR: 27798540.

Hainmueller, J. and D. J. Hopkins (2015). “The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes toward Immigrants”. In: *American Journal of Political Science* 59(3), pp. 529–548.

Halla, M., A. F. Wagner, et al. (2017). “Immigration and Voting for the Far Right”. In: *Journal of the European Economic Association* 15(6), pp. 1341–1385.

Hopkins, D. J. (2015). “The Upside of Accents: Language, Inter-group Difference, and Attitudes toward Immigration”. In: *British Journal of Political Science* 45(3), pp. 531–557.

Jokisch, B. and J. Pribilsky (2002). “The Panic to Leave: Economic Crisis and the “New Emigration” from Ecuador”. In: *International Migration* 40(4), pp. 75–102.

Mayda, A. M., G. Peri, et al. (2022). “The Political Impact of Immigration: Evidence from the United States”. In: *American Economic Journal: Applied Economics* 14(1), pp. 358–389.

Mendez, I. and I. M. Cutillas (2014). “Has Immigration Affected Spanish Presidential Elections Results?” In: *Journal of Population Economics* 27(1), pp. 135–171.

Moriconi, S., G. Peri, et al. (2022). “Skill of the Immigrants and Vote of the Natives: Immigration and Nationalism in European Elections 2007–2016”. In: *European Economic Review* 141, p. 103986.

Mudde, C. (2004). “The Populist Zeitgeist”. In: *Government and Opposition* 39(4), pp. 541–563.

Otto, A. H. and M. F. Steinhardt (2014). “Immigration and Election Outcomes — Evidence from City Districts in Hamburg”. In: *Regional Science and Urban Economics* 45, pp. 67–79.

Pan, W.-F. (2023). “The Effect of Populism on High-Skilled Migration: Evidence from Inventors”. In: *European Journal of Political Economy* 79, p. 102447.

Schmutz, B. and G. Verdugo (2023). “Do Elections Affect Immigration? Evidence from French Municipalities”. In: *Journal of Public Economics* 218, p. 104803.

Standaert, S. and G. Rayp (2022). *Where Did They Come From, Where Did They Go? Bridging the Gaps in Migration Data*. UNU-CRIS Working Paper Series.

World Bank (n.d.). *World Development Indicators*.

Xu, Y. (2017). “Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models”. In: *Political Analysis* 25(1), pp. 57–76.

Xu, Y. and L. Liu (2022). *Gsynth: Generalized Synthetic Control Method*.

Chapter 3

Migrant Voices

Abstract

Migrants can lead to profound economic and social changes in their home communities through the information they share with family and friends left behind. This paper leverages unique historical data and artificial intelligence to analyze the content of migrants' communications, the factors shaping them, and their potential effects at origin. We construct a novel dataset of over 6,000 letters from Irish emigrants in North America in the 19th and early 20th centuries. We characterize letter writers and their local environment using data from passenger lists and population censuses at destination and origin. Relying on Large Language Models (LLMs), we identify and classify topics in the letters. We first provide some novel descriptive evidence about migrants' communications. We then explore the data and find that some individual and local characteristics (e.g., gender, religion, the size of the Irish diaspora, and average incomes) significantly affect the presence and salience of topics related to economics, religion, and politics. We plan on expanding the analysis to explore how a large institutional change at destination (i.e., the expansion of Catholic churches across the US) may have affected social remittances and, in turn, religious outcomes in Ireland.

3.1 Introduction

International migration can be a powerful driver of economic and social change, transforming both host countries and the communities that migrants leave behind. Beyond trade and monetary remittances, migration can also facilitate the transfer of cultural beliefs, practices, and preferences that migrants acquire abroad and share with their origin communities —often referred to as social remittances (Levitt, 1998). There is evidence of social remittances on a wide range of dimensions, such as democratic values (Spilimbergo, 2009; Docquier et al., 2016; Barsbai et al., 2017), fertility preferences (Beine et al., 2013; Bertoli et al., 2016; Melki et al., 2024), and even gender norms (Tuccio and Wahba, 2018).[25] While it is often assumed that migrants will adopt and transmit the values of their host societies, the transmission of social remittances responds to complex factors, can be context specific, and thus is far from uniform.

Whether and how migrants diffuse social norms can depend on their traits and experiences, the socio-economic environments they encounter abroad, and even the conditions in their home communities. In some cases, migrants may reject foreign values and reinforce their own norms and identity. This backlash can be shaped by exposure to discrimination, exclusionary policies, or cultural differences (Fouka, 2019; Fouka, 2020; Jaschke et al., 2022). Similarly, different groups of migrants —e.g., by gender, social status, or religion— may respond differently to the same social environment, affecting which values, if any, they transfer back home. Despite the importance of understanding these questions, empirical evidence on how social remittances are transmitted is very limited, largely due to the lack of direct data on migrant communications and the nature of these exchanges.

This paper exploits novel data and a unique setting to analyze the content of migrant communications, the factors shaping them, and their potential consequences at migrant origin communities. Specifically, we gather over 6,000 letters from Irish migrants in North America during the 19th and 20th centuries, which we combine with multiple other sources. We exploit artificial intelligence methods (i.e., Large Language Models - LLMs) to analyze their content and assess how various migrant and local characteristics influenced social remittances. Additionally, focusing on the specific case of religiosity, we plan to analyze how a drastic change at destination (i.e., the expansion of the Catholic Church in the US) altered the content of the letters, and in turn, affected people's behaviors in migrant hometowns.

The Irish diaspora in North America is an ideal context for several reasons. First, since the Potato Famine in the late 1840s, Irish started migrating massively, leading to the formation of a sizable community abroad. Between 1850 and 1870, they were the largest migrant group

[25] See Levitt and Lamba-Nieves (2011) and Tuccio and Wahba (2020) for a review of the literature.

in the US, representing over 4% of the population. Second, literacy rates in Ireland were extremely high at the time, especially compared to other European countries, with close to 90% of the adult population being literate in 1901. These high literacy rates are crucial, as they imply that the vast majority of migrants were able to maintain correspondence with their families left behind. Third, Ireland was a largely Catholic country, yet with a significant share of Protestants in the North. This religious divide, in combination with the predominantly Protestant societies at destination, adds a layer of complexity and diversity that can influence religious identities and migrant communications. Finally, given the relevance of the Irish diaspora and the strength of its community, thousands of letters have survived the perils of time and been collected by researchers and other institutions over time.

We leverage data from multiple sources to construct a novel dataset containing thousands of migrant letters, which we then combine with migrant characteristics and socio-demographic information at both destination and origin. We gather most of the letters from two online repositories (i.e., the Irish Emigration Database and the Irish Emigrant Letters and Memoirs from North America), digitizing, harmonizing, and geolocating them. Using the sender's full name and the year of the letter, we implement a linking procedure to trace a sub-sample of migrants in passenger lists and retrieve basic information of letter writers (e.g., age, occupation, and year of arrival). Using micro-data from the population censuses of the United States and Canada during 1850-1920, we construct county-level measures to characterize the local environment for each letter (e.g., urbanization, share of foreign-born, share of Irish, average income, etc.). We also exploit the 1901 and 1911 censuses of Ireland to construct measures of religiosity based on names, which we apply to letter writers and the (de-anonymized) census data at destination to proxy religiosity at the local level.

We perform extensive tests to validate the quality of the letters and to ensure that the various sub-samples are representative (e.g., geolocated letters and those with the sender linked to the passenger lists). For instance, we show that there is a strong overlap in the geographical distribution of the letters sent across US counties and the distribution of Irish-born. We observe a similar pattern in Ireland, with a strong relationship between emigration rates and the distribution of letters received across counties. We find that missing information (e.g., sender or receiver's name, receiver's location) are mostly uncorrelated with local socio-demographics or with basic letter characteristics (e.g., its length, sentiment, etc.). We also analyze the linking approach and find that the characteristics of the sub-sample of senders that we can confidently link resemble those of passengers as a whole.

Given the large amount of letters and the vast heterogeneity in their content, we categorize them into topics and themes using state-of-the-art artificial intelligence (AI). We rely on two of the most popular Large Large Language Models (LLMs), Gemini 1.5 and ChatGPT 4o.

We use Gemini to inspect all the letters and help us identify core topics (e.g., *climate*, *health*, *economics*, *religion*, *politics*, *emigration*, and *relationships*). By interacting with the AI assistant in several iterations, we further identify themes (i.e., more nuanced ideas within topics) and sub-themes (i.e., more nuanced ideas within themes). For instance, "*Religious practices in daily life*" would be a theme within the topic *religion*, and "*Church attendance*" or "*Bible readings*" sub-themes. We then rely on ChatGPT API to analyze each letter one by one and identify the topics and themes present. We perform a human classification of 170 letters and validate the results of ChatGPT obtaining an accuracy above 90% in most topics (with both Type 1 and Type 2 errors generally below 10%).

Equipped with this data, we conduct several analyses. First, we explore the topics and themes to uncover new findings on migrants' communication during this period. We find that nearly all letters feature economic mentions, discussing information about wages and cost of living in North America, economic prospects, or comparisons with their previous life. Health is the second most common topic, including updates on migrant's health, news on health or death of acquaintances, or inquiries about others. References to climate, geography, and the settling process are very common, offering insights into changed migration patterns. Religious references are also prevalent, sometimes only as customary salutations, but others describing religious practices or the landscape in North America. Strikingly, over one-third of the letters address politics or social issues either in North America or in Ireland, highlighting the relevance of these questions. Finally, social relationships are widespread, such as news about other acquaintances or relatives.

Next, relying on the sub-sample of writers that we linked to passenger lists, we examine how personal characteristics influence the content of letters. The most important individual factor is gender. While men and women have the same likelihood of mentioning economic and health topics, women's letter are significantly more likely to have religious references, and less likely to have political ones. Catholic migrants are also significantly more likely to discuss religious issues. When we examine the intensive margin, i.e., the number of keywords related to a given topic, we also find that women discuss much less intensely political and economic issues. The time spent in North America is significantly associated with fewer mentions to the core topics and fewer keywords. This is mostly due to letters becoming shorter as migrants spend more time far from their homeland.

We find that migrants' local environment sharply influences the content of the letters. The size of the Irish community in the county is the most relevant factor, decreasing the likelihood of mentions to health, religion, or political issues. This evidence can be interpreted as if interactions with the family left-behind and the local community are substitutes. The level of urbanization of the county is negatively associated with mentions to climate (which includes geographic elements in it) and religion. When looking at the intensive margin (i.e., the number of words), the large negative effect of the Irish local community is reaffirmed. We also observe that in areas with higher income on average, letters contain less content discussing economics, religion, and politics.

We plan on extending these preliminary analyses in multiple ways. First, using the destination census data we will construct richer measures of social integration (e.g., intermarriage, naturalization, naming patterns, etc.) and assess their relationship with the content of letters. We will exploit the religious content of names to measure the religious landscape at the local level, calculate measures of diversity and polarization along religious lines, and explore the interaction of these with the letter writers potential affiliation. Finally, we will analyze the staggered expansion of Catholic churches across the US as a major sudden shock to religiosity, and examine its possible effects on (religious) social remittances, distinguishing both between Catholic and Protestant Irish and their local environment. In a final step, we will explore whether the potential effects on religious remittances can translate into changes in migrants' communities of origin. We will rely on the micro-data of the 1901 and 1911 Irish censuses to construct proxies of religiosity by cohorts using naming patterns and self-reported affiliation, among others.

Our paper contributes to several strands of literature. First, it contributes to the broader literature on migration and cultural change. Previous work has addressed various issues, including the assimilation of immigrants into the native culture (Abramitzky et al., 2020; Bandiera et al., 2019; Fouka et al., 2022), the selection of migrants along cultural traits (Docquier et al., 2020; Knudsen, 2024), or the diffusion of migrants' values through intergroup contact (Giuliano and Tabellini, 2020; Bazzi et al., 2023; Miho et al., 2024). Our work is closely related to Manacorda et al. (2024) who exploit cross-country variation to show that an increase in anti-migrant sentiments can lead to less democratic support in migrants' origin countries. We complement these studies by examining how the specific institutional and local socioeconomic conditions that migrants face can shape the content of social remittances overall and, consequently, influence cultural values and identity in their home communities.

Second, we make several contributions to the literature on social remittances. There is extensive evidence on the role of migrants in shaping political values at origin (Spilimbergo, 2009; Docquier et al., 2016; Manacorda et al., 2024), diffusing fertility preferences (Beine et al., 2013; Bertoli et al., 2016; Melki et al., 2024; Beach and Hanlon, 2023), as well as gender norms (Tuccio and Wahba, 2018; Lodigiani and Salomone, 2020; Diabate and Mesplé-Somps, 2019). These studies establish a clear link between destination values and changes at origin, but the actual causal pathways are often not directly observed or disentangled (e.g., the contribution of communications vs. returnees). Here, we focus exclusively on social remittances as measured directly in migrants' communications, which allows us to examine multiple social dimensions in a single setting and measure their intensity. Moreover, previous studies implicitly assume that transmission is homogeneous, i.e. that all migrants may adopt and transmit the norms observed at the destination in a similar way. We shed new light by analyzing how social remittances may vary across specific individuals.

Third, we contribute to the growing literature on the Age of Mass Migration, which has examined the short and long-term consequences of mass immigration in the United States (Abramitzky and Boustan, 2017; Tabellini, 2020; Sequeira et al., 2020), other countries in the Americas (Droller, 2018; Rocha et al., 2017), as well as on migrants' communities of origin (Karadja and Prawitz, 2019; Fernández, forthcoming). Our paper speaks to both strands, putting the emphasis on migrant's experiences abroad as a catalyst for social change at origin. The dataset of migrants letters that we have assembled is a contribution in itself, being a valuable resource to shed new light on variety of issues such as migrants' integration, feelings of discrimination, or attitudes towards other minorities.

Finally, our paper is related to a recent and growing literature exploiting text and image at scale as novel data sources to shed new light on major economic and social phenomena and historical events. This pioneering work examines a variety of issues, such as the persistence of narratives and their economic implications (Michalopoulos and Xue, 2021; Michalopoulos and Rauh, 2024), the process of cultural evolution and diffusion as manifested in hair-styles or newspapers (Voth and Yanagizawa-Drott, 2024; Posch, 2024), or the explorations of European settlers across Africa or Australia (Kerby et al., 2022; Kampanelis et al., 2023). In a related work, Michalopoulos et al. (2024) show that the experience of forced migration can lead to persistent changes in perceptions of discrimination and economic insecurity, which are transmitted over generations and reflected in the narratives of songs.

The paper is organized as follows. Section 3.2 provides a brief historical background. Section 3.3 describes the various data sources. Section 3.4 explains the processing of the data and its validation. Section 3.5 discusses the results, and section 3.6 concludes.

3.2 Background

3.2.1 Irish emigration, 1800-1920

Migration flows from Ireland to North America started to grow importantly in the early 19th century. Overall, it is estimated that about a million Irish left between 1815 and 1845, nearly three times as many as in the previous century (Connolly, 2022). While most of these early migrants were from the northern part of Ireland and hence Protestant, these migration waves expanded to the rest of the country soon making Catholics the largest group (Connolly, 2022).[26] In the 1840s, the Great Famine due to the potato blight (1846-1851) pushed hundreds of thousands of people into exile, marking the beginning of the Age of Mass Migration (Hatton, 2001). The size of the Irish diaspora in the United States nearly quadrupled between 1840 and 1850, and the one in Canada nearly doubled in size (Gráda, 2019). In the decades to follow, flows would continue to raise, making emigration a very standard choice for Irish people.[27]

During the second half of the 19th century, Irish were among the largest immigrant groups in the United States. They concentrated mostly at the end of the economic distribution. Labor competition was particularly fierce in the Famine years, as hundreds of thousands of migrants flooded American cities in a short period (Gráda and O'Rourke, 1997; Ferrie, 1997). Most of these migrants also had very few capital and relied on poor relief plans provided by the State (Hirota, 2016). Nevertheless, the gap in economic integration between pre- and post-Famine Irish migrants considerably reduced in the following decades (Collins and Zimran, 2019). By the mid-19th century, Irish immigrants had better economic prospects than American-born individuals themselves (Abramitzky et al., 2021b).

The combination of economic struggle and religious differences made Irish integration challenging. The 1850s saw the rise of the Know-Nothing party,[28] which was built on a strong nativist platform. Additionally, its members were staunchly anti-Catholics, fearing that the massive inflows of Irish could threaten the American institutions and establish a Papal state (Hirota, 2016). This open hostility towards Irish migrants took the shape of sporadic violence towards Catholics as well as deportation back to Ireland (Alsan et al., 2020; Hirota, 2016). These anti-Irish nativist positions continued for several decades until the end of the nineteenth century, when anti-migrant sentiment shifted towards those coming from southern and eastern Europe as well as Asia (Hirota, 2016).[29]

The relationship between Irish migrants and their homeland is not easy to unravel. On the one hand, migrants maintained contact with their relatives back home as they were sending letters and remittances (sometimes encouraging them to join them in America), and some were keeping themselves informed of the political issues between Ireland and

[26] Figure 3.F.1 in the Appendix shows the stark religious division between the north of Ireland, more Protestant, and the rest of country, overwhelmingly Catholic.

[27] It is estimated that close to million arrived to the US between 1860 and 1900. In 1876, the probability of migration for those entering the labour market was 50% (Fitzpatrick, 1980). Louis Paul-Dubois, visiting Ireland, wrote in 1908: Irish "*children are brought up with the idea of probably becoming emigrants*" (Paul-Dubois, 1908:p.389).

[28] While this party had branches in most of the States, it was most active in the Northeast of the United States.

[29] Hirota (2016): "The brutal violence Irish immigrants demonstrated during the Draft Riots of 1863 and Orange Riots of 1870 and 1871, both in New York City, reinforced the stereotype of the Irish as violent and ignorant people lacking the capacity for rational judgment, an indispensable quality for citizens in participatory democracy."

the rest of the United Kingdom. On the other hand, those who sent letters often expressed their homesickness and their regret of coming to America once they realized that life there (especially for discriminated low-skilled migrants) did not correspond to the American myth (Miller, 1985; Miller and Boling, 1991; Avila-Ledesma, 2019). As a consequence, many considered themselves involuntary exiles. Still, this negative information flow from the New World did not discourage young workers in Ireland to migrate as well, especially since monetary remittances enabled them to do so. Finally, many migrants gradually lost their ties with Ireland and considered the United States as their new home (Miller, 1985).

3.2.2 Migration and letter writing

The study of migrant letters has dramatically increased in the last decades as researchers gathered thousands of letters from various origins and time periods. An important literature focuses on Irish migrant letters as those have been extensively collected and made available by the work of Kerby Miller and others (Miller et al., 2003). Existing studies explore various aspects of letter writing. Focusing on Irish in Australia, Fitzpatrick (2006) shows for instance that letters usually follow a similar structure composed of introductory phrases, references to correspondence, references to health, affirmations of religious faith, and personal messages. This corroborates the idea that the primary purpose of letter writing is kin keeping, i.e., the "efforts expended on behalf of keeping family members in touch with one another" (Rosenthal, 1985; Hurlburt, 2017).

Beyond the structure of the text, several studies also explore the differences in content due to the writer's gender. For instance, Moreton (2015) focuses on the Lough collection in the Irish letters and uses the frequency of words co-occurrences and of specific structures in the text to infer the emotional bond between a letter writer and her family. Amador-Moreno (2016) focuses on a larger corpus of Irish migrants in Argentina, and finds that letters written by women express a closer psychological and emotional proximity with the letter recipient. This type of analysis focusing on similar but not identical words also uncovered differences in how men and women consider their homeland. Using letters from Irish migrants in several countries, Avila-Ledesma and Amador-Moreno (2016) find that men make a clear distinction between the (host) "country" in which they live, and their "home" (Ireland), emphasizing the fact that many migrants consider migration as an exile. This is not visible in letters written by women, who use both words interchangeably to speak about the United States or Ireland.

The analysis of migrant letters has also been made in various other countries and time periods. Some examples include Zempel (2000) on Norwegian letters, Munch (2021) on Danish letters, Barton (2012) on Swedish letters, and Alroey (2011) on Jewish letters. Gerber (2006) focuses on selection patterns among German letter writers and finds a positive selection in terms of education.[30] Several studies show that the tie between migrants and their relatives could sometimes extend beyond family as reading a letter from abroad could be a collective event for an entire village or neighbourhood.[31]

These studies use a large variety of methods. Some focus on a couple of individuals or families and follow their life through multiple letters (Eyford, 2015; Hurlburt, 2017; Dounia, 2020), while others use larger corpus to extract common patterns across letters (Borges, 2016), sometimes using a quantitative approach (Amador-Moreno, 2016; Avila-Ledesma, 2019; Moreton, 2015). A shared characteristic of most of these papers is the application of a bottom-up approach, starting from very specific case studies and expanding progressively to larger samples. We are not aware of any study of the scope and nature of the present work, combining thousands of letters with systematic socio-demographic data and analyzing their content empirically.

3.3 Data sources

3.3.1 Migrant letters

We gather thousands of letters from Irish migrants in the United States and Canada from two main sources: the Irish Emigration Database (IED) and the Irish Emigrant Letters and Memoirs from North America (IMIRCE). The IED was created in the 1980s by the Ulster American Folk Park in Northern Ireland. It provides access to various types of text documents, such as newspapers, diaries, and wills. More than 3,200 migrant letters from several public repositories in Belfast as well as private donations were gathered, transcribed, and made available on the IED website. The IMIRCE website, developed and maintained by the University of Galway, provides access to the collection of letters compiled by Prof. Kerby Miller since the 1970s. Prof. Miller devoted several decades to collecting Irish migrant letters from archives, libraries, and donations, transcribing them, and carrying out pioneering work (Miller, 1985). So far, more than 3,600 letters are available, and this number continues to grow as the digitization process continues.[32] Besides providing the original letters, the IED and

[30] Some studies also document that a side effect of migration was the increase in literacy rate, as letter writing was the only means of communication and it was costly to ask someone else to do it (Gabaccia, 2000:p.90)

[31] This could also be due to the fact that literacy rates were sometimes low, so only a handful of people were reading letters to everyone (Markelis, 2006). According to Borges (2016): "In rural and small-town Portugal, correspondence was not distributed to individual households, so public distribution of letters became a performative moment with both individual and collective meaning." Ishiguro (2011) in her study of British migrants in the Age of Mass Migration notes: "While letters were usually written by and addressed to individuals, there was a general understanding among middle-class Britons that family correspondence was to be shared more widely, read aloud to one another and even sent from town to town."

[32] Since the letters are uploaded as images rather than plain digital text, we perform Optical Character Recognition (OCR) using `tesseract` (Ooms, 2024) to extract the text.

IMIRCE websites display key information such as the year, full names of sender and receiver, and their respective locations. Importantly, letters are often classified in collections which relate to the families concerned or the primary sources.[33]

3.3.2 Passenger lists

We collect socio-demographic information on Irish migrants to the United States and Canada from passenger lists for the years 1820-1924. The US data originate from the Castle Garden (1820-1891) and Ellis Island (1892-1924) databases, while the Canadian data come from ship manifests of official ports (1865-1922).[34] We look for passengers with the exact same name (or extremely similar) as the writers in our letters database and retrieve information on the year of arrival, age, and occupation.

3.3.3 Census data

Ireland. We exploit the 1901 and 1911 full micro-censuses of Ireland to construct several measures across counties and birth cohorts. These include literacy rates, the share in various occupations, the share of women working, and the share of Catholics and Protestants. We exploit the self-reported information on religion to identify names distinctively associated with a given religion (see section 3.F in the Appendix).

United States & Canada. We use micro-data from the population censuses of the United States and Canada (1850-1920) to construct various socio-demographic measures at the local level (Ruggles et al., 2024). In particular, for every decade and county,[35] we compute the share of the population living in cities, the average income score (distinguishing natives, foreign-born, and Irish-born), and the share of the adult population born abroad and in Ireland. As religion is not observable in US censuses, we compute the share of Irish holding names strongly associated with a particular religion, using the same method as mentioned in the previous paragraph.

3.3.4 Other sources

We use data from the UK Data statistics across Irish counties over the period 1840-1920 (Clarkson et al., 1997) to construct several measures of emigration rates and socio-economic

[33] Table 3.A.1 in the Appendix provides more information on the primary sources. See IED and IMIRCE sites.

[34] We obtain the data from the websites of Ellis Island, FamilySearch , Dunbrody, and Ancestry. For the US, the data cover the port of New York for 1820-1891, as well as Boston, Baltimore, New Orleans, and Philadelphia for the famine years (1846-1851). For Canada, the sources cover all main ports (i.e., Quebec, Montreal, Halifax, Saint John, Vancouver, Victoria, and Sidney). Information prior to 1865 is largely lost.

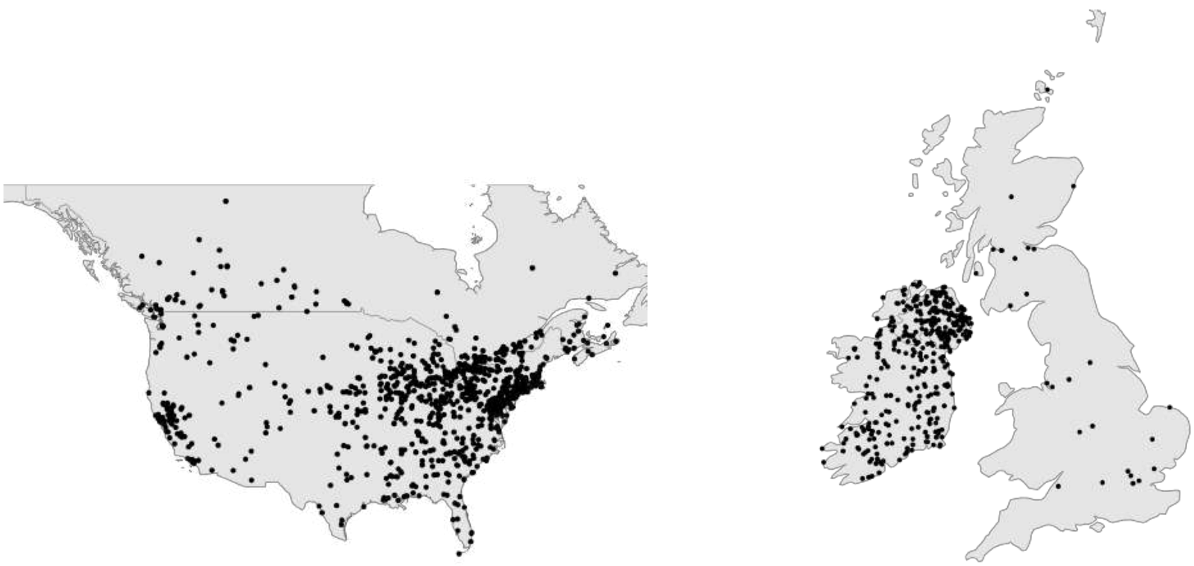
[35] We rely on US counties and Canadian provinces as main geographical units. The Canadian censuses of the 19th century do not provide a more granular (harmonized) geographical variable.

indicators, such as the share of urban population or employment rates.

3.4 Data construction & validation

3.4.1 Georeferencing letters

To match the locations of writers and receivers with counties in the US, Canada, and Ireland, we first standardize all the available information and then geocode it using ArcGIS.[36] We ensure that all name variations (e.g., 'Nw York', 'N.Y.C.', 'N. York') are harmonized and whenever a location could ambiguously refer to both a city and a state/province (e.g., New Work, Quebec) we assume it refers to the city. Tables 3.A.2 and 3.A.3 in the Appendix describe the relative importance of each country (and level of precision) in the entire dataset. The precise sender location is available in 95.8% of the letters sent from the US and Canada.



(**a**) Letter senders in North America (**b**) Letter receivers in the United Kingdom

Figure 3.4.1: Location of letter senders and receivers

Notes: The figures display the location of senders and receivers among the letters sent from North America to the United Kingdom during 1840-1930.

Among the receivers, we have missing information on the precise location for 25%[37]. To partially address this issue, we exploit the fact that letters are not independent from one another but belong to collections and that a single migrant often sent multiple letters. If a sender

[36] We also use Google's Geocoding API to verify that the geolocations are identical. We perform some robustness checks excluding from the sample the few cases in which ArcGIS and Google provide different results.

[37] In 20% of the cases we don't know if the receiver is in Ireland, even though this is very likely.

(i.e., letters from the same collection, same sender name and location, and temporally close) writes letters to people in a unique location, we assume that any letter without information was sent to that location too. Using this approach, we increase the share of senders with precised location from 50% to 53.4%.[38] Figure 3.4.1 shows the exact location of letter writers and receivers. The large majority of letter writers are located in the Northeast part of the United States, and in the Southeast part of Canada, in Ontario. Receivers are distributed all over Ireland, with a larger concentration in the Northern part (in part due to our sources over-representing the North).

3.4.2 Linking letters to passengers

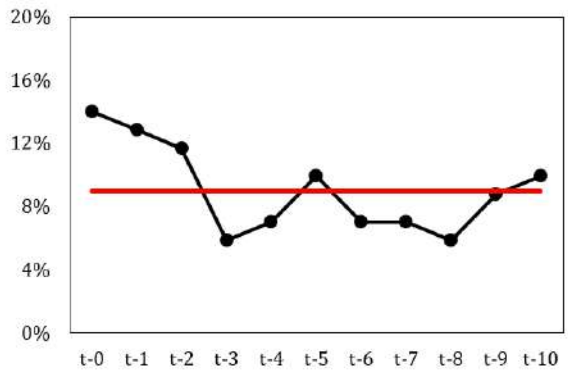
We leverage the information from passenger lists to infer the socio-demographics of letter writers. Contrarily to the linking of other data sources, the limited information makes this task challenging. To motivate our approach, we start by examining the number of potential links (passenger with the exact same name as the writer) in a window of 10 years before the letter.

Our linking procedure relies on two characteristics of letter writers, namely their full names and the year in which they sent their first letter. We start by computing the name similarity between letter writers and passengers,[39] and discard pairs whose similarity is lower than 0.95.[40] The second step uses the fact that the first letter is more likely to be written in the first years following the arrival. Indeed, when a letter writer has more than two potential links, those potential matches are quite evenly distributed in the 10-year period (see Figure 3.4.2b). However, in the case of letter writers with at most two potential links (in a 10-year window), there is much higher probability that those potential links actually lied in the three years preceding the letter (see Figure 3.4.2a). Therefore, we start by searching for passengers that have a similar name and arrived in the three years before the letter was written. If we find at least one link in those three years, we stop searching for this writer. If we did not find any link, then we look for passengers with a similar names in the three years before. We do this "incremental" procedure until we reach an arbitrary threshold of 21 years. For the sake of comparison, we also apply a "naive" procedure in which we use the same values for all letter writers. The difference with the "incremental" procedure is that for each 3-year period, we consider all letter writers and not only those for whom we haven't found a link yet. Figures 3.D.1 and 3.D.2 in the Appendix show the outcomes of the linking procedure, using

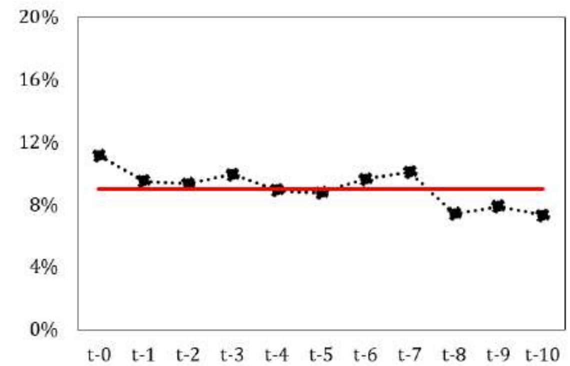
[38] Section 3.C adds more details on the cleaning and geolocating processes. Section 3.C-II provides more information on the various strategies to infer the location of letter receivers.

[39] We use the Jaro-Winkler measure which is frequently used in the literature on census linking since it was designed to accommodate typos in full names (Jaro, 1989; Abramitzky et al., 2021a).

[40] This is stricter than the algorithm usually applied in census linking which uses a similarity of 0.9 or higher (Abramitzky et al., 2021a).



(a) Distribution of links (if 2 or fewer)



(b) Distribution of links (if more than 2)

Figure 3.4.2: Distribution of potential links by year relative to the letter

Notes: The figure shows the distribution of potential links for letter writers in the decade prior to the letter (i.e., individuals in the passenger list with the same full name). Given the 11-year period, the expected links per year are 9% of the total. Figure (a) shows that when there are only 2 or fewer passengers with the same name as the writer, the likelihood of observing them in the first years before the letter is significantly higher.

respectively the ”naive” and the ”incremental” approach.

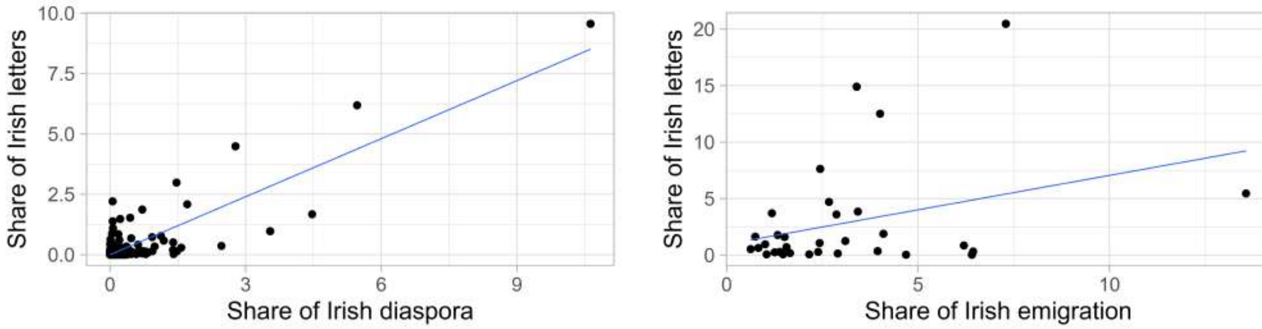
In many cases, there are more than three potential links for a letter writer, making it impossible to confidently attribute their personal characteristics to the letter writer. We found a unique link in about 21% of the cases, and we found at most two links in about 8% of the cases.[41] Overall, we end up with a linking rate of about 29%, which is close to other record linkage settings (Abramitzky et al., 2021a).

3.4.3 Data validation

We perform a number of analyses which support the quality of the data and suggest that it may be representative of the Irish diaspora in the period studied. First, the timing of the letters overlaps well with the arrival of Irish migrants to North America as shown in Figure 3.B.1 in the Appendix. In other words, the data provides a meaningful representation of the entire mass migration era. Secondly, we explore the geographical distribution of the letters and find that it closely matches other sources at both destination and origin. As shown in Figure 3.4.3a, the share of letters sent across US counties (1840-1920) is strongly correlated with that of Irish immigrants (1880). Similarly, Figure 3.4.3b shows a strong relationship between the share of letters received across Irish counties (1840-1920) and emigration rates (1840-1900).[42]

[41] When we have two links, we take the average of their characteristics. Results using only the characteristics of letter writers for whom there is a unique link are very similar.

[42] Note that using alternative years for the distribution of Irish across US counties or that of emigration rates across Irish counties yield a very similar picture.



(**a**) Irish across US counties, 1880

(**b**) Emigrants across Irish counties, 1850-1900

Figure 3.4.3: Correlation between letters location and Irish migration data

Notes: Figure 3.4.3a shows the correlation between the distribution of Irish migrants in the United States in 1880 and the share of migrant letters sent across US counties (1840-1920). Figure 3.4.3b shows the correlation between the share of Irish emigration 1850-1900 and the share of migrant letters received across Irish counties (1840-1920).

We assess whether the availability of certain information (i.e., sender name, receiver name and location) is associated with characteristics of the sender's location such as the level of economic development or the size of the Irish community. As shown in Table 3.A.4 in the Appendix, we find no evidence that county characteristics are correlated with missing information on the sender's name. This evidence is crucial for the linking of senders and passengers, as it points to no selection in this dimension. We see no relationship either with the size of the Irish community. The average income of the county is however negatively associated with having missing information about the receiver's name and location.

Given that we can only reliably link 29% of the senders to the passenger lists, we investigate whether letter writers are representative of Irish immigrants overall by comparing the socio-demographic information of both groups as illustrated in Table 3.4.1.[43] Overall, we find that the average writer and passenger are very similar in terms of age and socio-economic status. Men are slightly over-represented in our sample of letter writers, as well as individuals with a more Protestant name, likely the result of the IED source over-representing northern Ireland.

3.4.4 Content classification

We categorize the content of letters at various levels of specificity by identifying broad *topics*, *themes*, and *keywords*. *Topics* refer to general categories such as health and religion. *Themes* represent recurrent ideas or narratives that provide a more nuanced view on specific topics. For example, a common health-related theme is the "*reaction to the death of a relative*". Finally, for each of the topics we collect a comprehensive list of keywords associated with them. While

[43] Separate results for IED and IMIRCE are available in tables 3.A.5 and 3.A.6 in the Appendix.

Table 3.4.1: Balance table between all passengers and linked letter writers.

	All (N=1831746)		Linked (N=257)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
Age (at arrival)	26.33	9.38	24.70	9.90	-1.63	0.009
Female	0.45	0.50	0.26	0.44	-0.18	$<$0.001
Arrival in New York	0.95	0.22	0.96	0.18	0.02	0.140
Occupational income score (OIS)	10.40	5.18	12.29	7.46	1.89	0.002
Socio-eco. status index (SEI)	12.64	9.43	12.52	12.30	-0.12	0.905
Earnings index score (ERS)	9.26	15.06	14.57	22.27	5.31	0.003
Educational index score (EDS)	4.19	5.60	4.51	6.32	0.32	0.521
Catholic name	0.22	0.41	0.09	0.29	-0.12	$<$0.001
Protestant name	0.05	0.22	0.20	0.40	0.15	$<$0.001
Neutral/Unknown name	0.73	0.44	0.70	0.46	-0.03	0.315

Female is a dummy variable indicating whether the individual is a woman. Arrival in New York is a dummy indicating whether the individual's port of arrival is New York. OIS, SEI, ERS, and EDS are scores associated with occupation names as listed in the HISCO database. For religiosity, we use our measure based on name frequency in Irish District Electoral Divisions.

the topics and themes will measure the presence of content in a binary way (yes/no), the keywords will be useful to measure the salience of specific issues.

Given the vast diversity in the content of letters, we rely on cutting-edge Large Language Models (ChatGPT version 4o-mini and Gemini 1.5 Pro) to identify *topics* and *themes* in the whole corpus and then categorize each letter separately.

To identify the *topics* in a non-deterministic way, we upload the whole set of letters in NotebookLM (an AI research assistant based on Gemini 1.5 Pro) and prompt the artificial intelligence as follows:

> "*Here are hundreds of letters from Irish migrants in North America during the 19th and early 20th century, writing to family and friends left behind. Based on your reading of all the material, I would like you to identify the most important topics featured in the letters (e.g., health, economics, etc.).*"

This prompt tends to output between five and seven topics. We repeat the procedure in independent conversations ten times, which results in a total of seven recurrent topics: climate, health, economics, politics, religion, migration, and relationships. We will on focus on six key topics (i.e., climate, health, economics, politics, religion, and migration) as both core outcomes and benchmarks for assessing the accuracy of our classification. The topic "relationships" is ubiquitous, being present in almost all the letters.

After having identified the topics, we interact further with the artificial intelligence to obtain the main themes. As before, we use a very generic prompt:

> *"Now, I would like you go more in detail and make a long list of the most popular themes within the topic of "X". By themes I mean recurrent ideas or narratives that provide a nuanced view of that topic. I want you to review all the letters and be comprehensive. Rather than long explanations, I want you to identify as many themes and sub-themes as you consider appropriate."*

Since the artificial intelligence provides a different classification in each conversation, we repeat this procedure five times. We then combine all the output into a single document, feed it to the AI instead of the actual letters, and request to identify the themes one last time. By doing this, all redundant themes are combined and we obtain a comprehensive classification. The full list of themes and sub-themes is displayed in the Appendix, Section 3.E.

To categorize each letter as featuring a given topic, we prompt ChatGPT with a very generic text. To reduce the risk of mixing information from different letters, we exploit the ChatGPT API and start one unique conversation for every letter with only one prompt:

> *"I will give you a letter written by an Irish in North America in the 19th or early 20th century. I want you to determine if this letter talks about the following topics: climate, health, economics, religion, politics, migration, relationships. For each topic, return a string containing the topic, followed by 1 if it is mentioned and some keywords (not sentences) that support your answer, separated by a comma. If the topic is not mentioned, return the topic name, followed by 0 as a string followed by "nothing". I want you to return a list with your answer for each topic. The order of topics in your answer MUST be climate, health, economics, religion, politics, migration, relationships. Here is the letter: <letter content> "*

To evaluate the performance of our AI-based classification approach, we create a validation dataset by manually classifying a random sample of 170 letters. We take a comprehensive approach, coding even minor references to a topic (e.g., "*Thank **god** you are well*" → religion, "*The **riots** have intensified*" → politics, "*This **winter** has been mild*" → climate). We also extract all the relevant keywords (e.g., god, riots, winter) to conduct further tests on the AI classification.

Table 3.4.2 shows the results of ChatGPT classification for our manually classified sample of letters. Overall, ChatGPT is able to recognize that a letter talks about multiple topics at once and gives very similar results to the human classification. The overlap is extremely high, around 90% for most topics. Both Type I and Type II errors are generally below 10%. Type I errors are higher in the case of climate (0.174) and politics (0.197). We examine in detail these false positives and find out that in the case of climate, they mostly arise from references

Table 3.4.2: Performance of ChatGPT on topic classification on our sample of 170 letters.

Theme	Our classif.	ChatGPT classif.	Same classif.	Type 1	Type 2
weather	0.506	0.554	0.855	0.174	0.095
health	0.759	0.807	0.916	0.082	0.024
economics	0.819	0.801	0.934	0.030	0.051
politics	0.325	0.367	0.898	0.197	0.093
religion	0.572	0.482	0.898	0.013	0.168
migration	0.382	0.705	0.614	0.496	0.079

Column 2 shows the share of letters that we classified as talking about a topic, and column 3 shows the share the ChatGPT classified as such. Column 4 shows the share of letters where ChatGPT provided the same classification as we did. Column 5 shows the type 1 errors, i.e. among all letters that were classified by ChatGPT as mentioning a theme, what is the share that we didn't classify as mentioning this theme. Column 6 shows the type 2 errors, i.e. among all letters that we classified as mentioning a theme, what is the share that ChatGPT didn't classify as mentioning this theme.

to seasons as a temporal marker (e.g., "Next Summer I will rebuild the house") rather than providing any climate-related information. In the case of politics, many of the misclassified letters discuss family issues related to inheritances, and hence contain mentions to juries and tribunals that ChatGPT mistakenly associates with politics.

Finally, we also compute a measure of sentiment for each letter using the so-called "Bing" dataset (Hu and Liu, 2004). More precisely, we compute the number of positive and negative keywords in each letter, as well as single metric in which we standardize the difference between the two.

3.5 Results

3.5.1 Descriptive evidence

We start by providing some descriptive evidence on the content of the letters. We explore six core topics: climate, health, economics, religion, politics, and migration. We restrict our attention to the sample of letters sent by Irish migrants in North America (US & Canada) to a receiver in Ireland.

Figures 3.5.1 and 3.5.2 illustrate how the presence of a topic in letters (i.e., extensive margin) and the number of keywords associated to those topics (i.e., intensive margin) evolved over time. Some topics, such as economics, appear in almost all letters, no matter the decade of interest. This is expected as Irish migrated to escape the poor conditions of their home country, and therefore frequently gave news of their economic integration in North America. Yet, over time, economics took less and less place in the letters and the number of keywords associated to this topic decreased by more than 50% in 60 years. Other topics too, such as migration, became gradually less present over the second half of the 19th century. This is understandable given that the migratory pressures of the 1840s and 1850s gradually faded away. Therefore, when the number of Irish migrants started to decrease in the second half of the 19th century, references to migration in letters also became less frequent. The presence of religion and politics varied importantly during this period. Interestingly, we can see that after a slowdown in the 1870s and 1880s, it increased substantially in the following decades. Part of the explanation for this could be the tense situation in Ireland in the early 20th century, eventually leading to the Easter Rising in 1916 and to the war of Independence at the end of the 1910s. The process of integration in the US and the cultural divide between Protestants and Catholics could have also intensified over time, especially as the Catholic institutions became more and more consolidated in North America.

We can also distinguish some patterns from letters in which the relationship between the writer and the receiver is known. Figure 3.B.4 in the Appendix shows that letter writers tend to write more frequently to their siblings, followed by their parents and their extended family (uncle and aunt). This is an interesting feature as previous studies have shown that migrants sometimes self-censor in their letters, especially when writing to their parents (Gerber, 2006).[44] In a future analysis, we will use this information to explore whether migrants address some issues with their siblings but not with other family members.

Finally, the number of years spent in the country affects some characteristics of the letters, particularly their length. As shown in Figure 3.5.3a, the length of the letters decreases decreases by 40% once migrants have spent over 15 years in North America. Migrants seem to lose contact with relatives in their homeland but not progressively, rather when a long time-span has passed. This could happen for several reasons. For instance, as migrants become more and more attached to their new country of residence, they may decide to eventually break ties with their home country. Another plausible explanation is that their contacts in Ireland joined them in North America. While the length of letters decrease over time spent abroad, the average sentiment of letters (i.e., positive vs. negative words) is overall quite stable, as shown in figure 3.5.3b. Additionally, the average sentiment is close to 0, indicating a certain neutrality as positive and negative news balance out.

[44] This is either because they don't want their family to worry too much, or because they left without the approval of their parents and therefore acknowledging their difficulties would cause embarrassment.

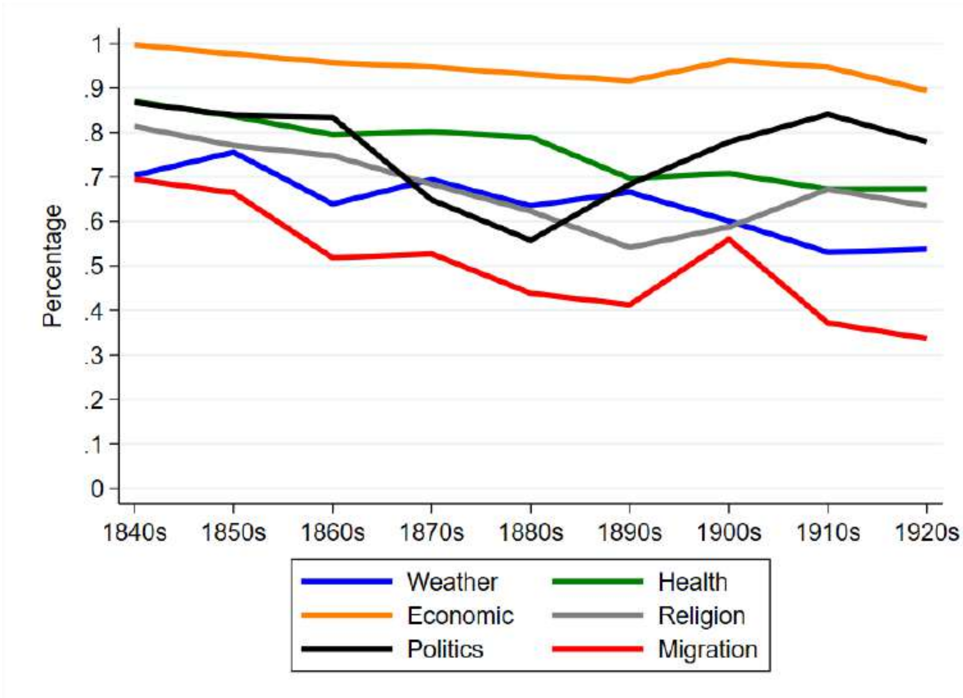


Figure 3.5.1: Average topic presence by decade, 1840s-1920s
Notes: The figure shows the share of letters mentioning every topic (at least one reference) by decade.

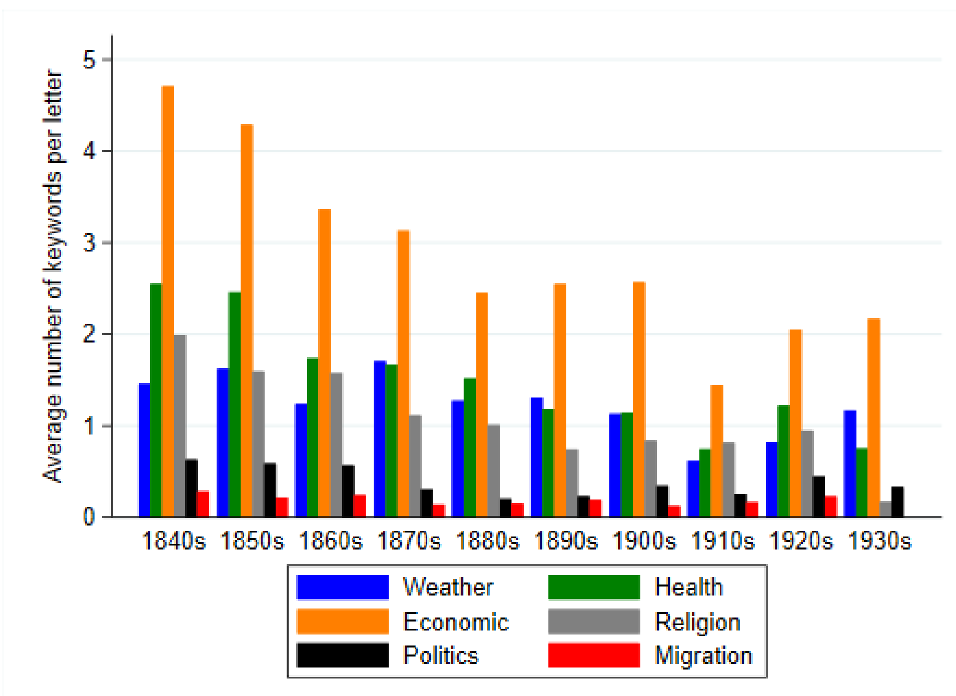
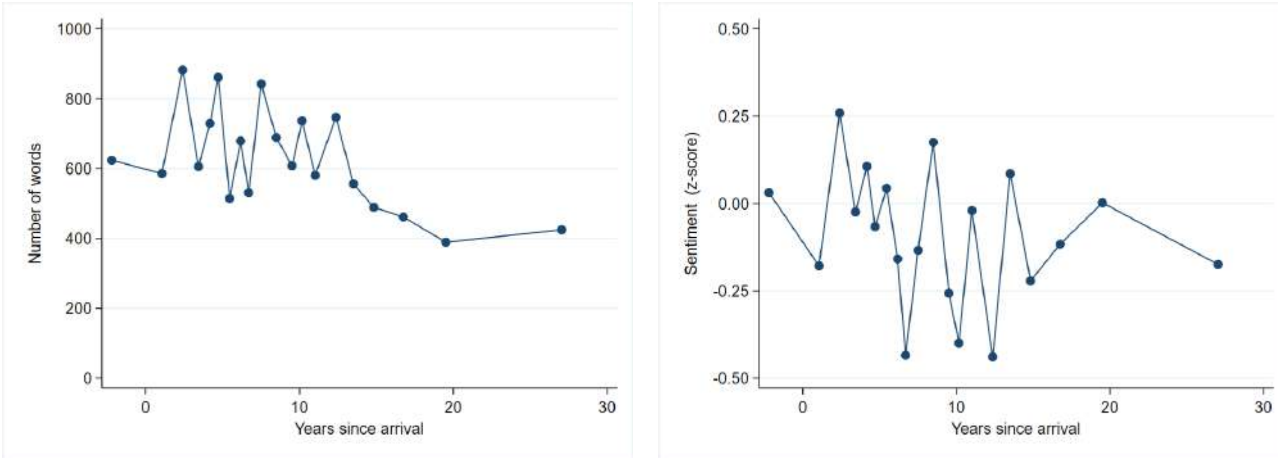


Figure 3.5.2: Average number of keywords by decade, 1840s-1920s
Notes: The figure shows the average number of keywords by topic and decade (including zeros).



(**a**) Word-count and years since arrival

(**b**) Sentiment and years since arrival

Figure 3.5.3: Years at destination and letter's size and sentiment

Notes: Figure 3.5.3a shows the average number of words in a letter by the years since arrival. Figure 3.5.3b shows the average sentiment in a letter (i.e., a standardized measure of positive and negative words) by the years since arrival. The figures rely on the sub-sample of letters by migrants confidently linked to one or two passengers.

3.5.2 Migrant characteristics

Do all migrants talk about the same topics and to the same extent? This section examines how migrants' socio-demographic characteristics shape the content of letters and the intensity of social remittances.

We explore the effect of those characteristics both on the extensive margin, i.e., whether a letter mentions a particular topic or not, and on the intensive margin, i.e., the number of keywords associated with a specific topic in a letter. We focus on the the sub-sample of letter writers that we can directly link to either one or two individuals in the passenger lists (in a horizon of one decade). Given that some of them wrote several letters, we have a final sample of 473 letters written by 257 individuals. We estimate the following equations by ordinary least squares (OLS)[45] :

$$TopicX_{ict} = \alpha_c + \alpha_t + X_i\Pi + u_{ict} \tag{3.1}$$

$$KeywordsX_{ict} = \alpha_c + \alpha_t + X_i\Pi + u_{ict} \tag{3.2}$$

where $TopicX_{ict}$ is a dummy variable indicating whether a given topic (X) is mentioned in letter i, and $KeywordsX_{ict}$ is the number of words related to that topic. The subscripts c and t refer to country and decades, respectively. The term X_i is a vector of demographic

[45] Given that the dependent variable in equation 3.1 is binary, we also use the probit method. Results, displayed in Table 3.A.7 in the Appendix, are very similar.

characteristics, including age, gender, religiosity (based on names), family relationship with the receiver, and years since arrival. The country fixed effects (α_c) account for time-invariant differences between the United States and Canada, while the decade fixed effects (α_t) control for unobserved factors that vary over time but similarly across space (e.g., more references to war in the decade of the American Civil War).

Table 3.5.1 displays the results of the estimation of equation 3.1 for all topics separately. We can extract several insights from this table. First, the gender of the migrant is a major determinant of the the content of the letters. Women are significantly more likely to talk about climate and religion, but less likely to discuss politics in their letters. The effects are sizable, as large as 14 percentage points gap. We also find that individuals with strongly Catholic names are much more likely (than those with neutral names) to mention religion in their letters, an effect that is not observed for writers with a Protestant name. This could be due to the fact that the cultural clash was stronger for Catholic migrants, who arrived in a country with a Protestant majority and who were sometimes discriminated against because of their faith.

The number of years since arrival tends to have a negative effect on the likelihood that a topic is mentioned in the letter. This is mostly due to the fact that a longer period at destination is correlated with shorter letters, as previously discussed. Once we condition on the length of the letter (Table 3.A.8 in the Appendix), this negative effect of years since arrival vanishes for almost all topics.

While the extensive margin is interesting in itself, the differences between a letter that talks about a topic and one that does not can be thin[46]. The same applies in reverse, two letters can both mention economic issues, but one may devote extensive content to it while the contain a minor reference. Table 3.5.2 displays the results of the estimation of equation 3.2 for all topics[47]. Combining the two tables gives more insights on other aspects. In particular, table 3.5.1 shows that women are not less likely to mention economics in general in their letters, yet when they do so, they give it much less importance than men. We observe the opposite pattern for health-related topics, which appear equally in letters written by men and women but are detailed in slightly more length in those written by women. Interestingly, the role of religious names is reversed when we look at the intensive margin. While Catholic names increased the probability of mentioning religious topics, they do not play any role on the number of keywords associated to religiosity. On the opposite, Protestant names do not affect the likelihood of talking about religion, but when this topic is in a letter, they increase the number of keywords associated to it. Finally, the negative effect of the length of stay is also

[46] For instance a single mention of "God" would suffice for a letter to be classified as mentioning "religion".

[47] Results with increasing number of controls and fixed effects for topics "religion" and "economics", at the extensive and intensive margins, and at the individual and county-level, are available in tables 3.A.9-3.A.16 in the Appendix.

Table 3.5.1: Relationship between migrant socio-demographics and topic presence

	(1) Climate	(2) Health	(3) Economic	(4) Religion	(5) Politics	(6) Migration
Female	0.125**	0.048	-0.008	0.140**	-0.140**	-0.029
	(0.059)	(0.052)	(0.031)	(0.055)	(0.062)	(0.061)
Age	-0.005**	0.002	0.001	-0.001	0.003	-0.005**
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)
Catholic name	-0.065	-0.102	-0.027	0.162*	0.076	-0.083
	(0.093)	(0.086)	(0.057)	(0.087)	(0.090)	(0.092)
Protestant name	0.200***	0.022	0.030	-0.015	0.055	0.067
	(0.067)	(0.066)	(0.033)	(0.070)	(0.075)	(0.081)
Years since arrival	0.000	-0.007**	-0.006**	-0.007*	-0.012***	-0.006
	(0.004)	(0.003)	(0.002)	(0.004)	(0.004)	(0.004)
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.64	0.78	0.94	0.68	0.62	0.47
R^2	0.07	0.11	0.03	0.09	0.07	0.08
Observations	473	473	473	473	473	473

Notes: The sample comprises all the letters sent from North America to Ireland during 1840-1929 with a linked letter writer. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

present at the intensive margin.

Overall, there is a lot of heterogeneity in the role of individual characteristics on the content of letters. The gender of the writer appears to be one of the main determinants, which corroborates some findings in the existing literature (Avila-Ledesma and Amador-Moreno, 2016). The religion of the writer has an asymmetric influence depending on the affiliation and the the margin considered, results that deserve further analysis.

Table 3.5.2: Relationship between migrant socio-demographics and topic intensity (keywords)

	(1) Climate	(2) Health	(3) Economic	(4) Religion	(5) Politics	(6) Migration
Female	-0.180	0.586	-4.389***	-0.146	-2.556***	-0.441***
	(0.366)	(0.375)	(1.089)	(0.422)	(0.648)	(0.151)
Age	-0.030*	0.003	-0.028	0.025	0.021	-0.017**
	(0.016)	(0.015)	(0.045)	(0.028)	(0.051)	(0.007)
Catholic name	-0.615	0.247	1.840	0.177	2.444	-0.461**
	(0.450)	(0.518)	(1.694)	(0.492)	(3.204)	(0.183)
Protestant name	1.480**	0.358	3.347*	1.784*	1.643	-0.264
	(0.627)	(0.455)	(1.851)	(1.043)	(1.747)	(0.178)
Years since arrival	0.018	-0.022	-0.176**	-0.094***	-0.143***	-0.017*
	(0.023)	(0.021)	(0.069)	(0.032)	(0.048)	(0.009)
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	2.73	2.72	10.43	2.79	3.71	0.98
R^2	0.10	0.06	0.14	0.08	0.08	0.08
Observations	473	473	473	473	473	473

Notes: The sample comprises all the letters sent from North America to Ireland during 1840-1929 with a linked letter writer. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.5.3 Local socio-economic environment

How does the socio-economic environment in migrants' host communities affect their communications? To shed light on this question, we use our extended sample of letters sent from North America to Ireland which contain information on the location of the sender. We estimate a multivariate regression with various local-level characteristics:

$$TopicX_{ilct} = \theta_c + \theta_t + W_{lct}\Gamma + v_{ilct} \tag{3.3}$$

$$KeywordsX_{ilct} = \theta_c + \theta_t + W_{lct}\Gamma + v_{ilct} \tag{3.4}$$

where $TopicX_{ilct}$ is a dummy variable indicating whether a given topic (X) is mentioned in letter i and $KeywordsX_{ilct}$ the actual number of words associated with that topic. The subscript l refer to the local area (i.e., counties in the US, provinces in Canada), c to country, and t to decade. The vector W_{lct} gathers a host of local characteristics that vary over time, namely urbanization rates, the share of foreigners, the share of Irish-born, the average income score, and the average income gap between natives and Irish-born.

We focus first on the results on the extensive margin, displayed in Table 3.5.3. The most prominent factor is the size of the Irish diaspora in the county. A larger local network is negatively associated with the pressence of almost all the topics of interested. This could be due to the fact that migrants tend to substitute their network back in Ireland with their local network, leading to less connections with their homeland. The effects are large and significant. For instance, increasing the share of Irish born by 10 percentage points, would be associated with a decrease in the likelihood of mentioning religion or politics in the letter by about 4.9 and 4.3 percentage points respectively. We see that urbanization plays an important role. Letters written in more urban areas have less mentions of climate, which is consistent with the fact that many climate-related themes are linked to agriculture. Migrants in urban areas are also less likely to mention religion in their letters, which could be an outcome of being exposed to a much liberal environment. Finally, the income gap between Irish migrants and natives also plays a noticeable role. In particular, it increases the likelihood that a letter talks about religion and migration. One way to interpret this finding is that places with a large gap indicate areas where Irish migrants faced more discrimination. This discrimination could go well beyond the labor market, and be connected to their faith too. In such a context, migrants might emphasize how their religion is a source of attack, and at the same time, discourage other migrants from making the move to North America.

Table 3.5.3: Migrant socio-economic environment and topic presence

	(1) Climate	(2) Health	(3) Economic	(4) Religion	(5) Politics	(6) Migration
Urbanization (%)	-0.108**	-0.047	0.009	-0.105***	0.058	0.027
	(0.044)	(0.038)	(0.018)	(0.035)	(0.040)	(0.042)
Foreign-born (%)	0.236**	0.058	-0.072	-0.095	-0.012	0.099
	(0.114)	(0.097)	(0.057)	(0.109)	(0.102)	(0.117)
Irish-born (%)	-0.982***	-0.694***	-0.143	-0.490*	-0.434*	-0.200
	(0.274)	(0.243)	(0.130)	(0.250)	(0.248)	(0.271)
Avg. income (std.)	0.016	0.028**	0.011*	0.018	-0.011	-0.003
	(0.015)	(0.013)	(0.007)	(0.014)	(0.014)	(0.016)
Income gap native-Irish (std.)	0.048***	0.047***	0.001	0.077***	-0.018	0.050***
	(0.018)	(0.018)	(0.008)	(0.018)	(0.017)	(0.019)
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.65	0.75	0.94	0.64	0.70	0.49
R^2	0.03	0.02	0.01	0.05	0.09	0.04
Observations	2,659	2,659	2,659	2,659	2,659	2,659

Notes: The sample comprises all the letters sent from North America to Ireland during 1840-1929 with a precised location. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Turning to the results on the intensive margin shown in table 3.5.4, we see that some variables that did not affect the likelihood of a topic being present in a letter now have a significant effect on the extent it may be discussed. For instance, letters written in counties with a higher average income give less importance to multiple topics, such as economics, religion, and politics. This effect is largely driven by migrants settling in richer areas writing shorter letters. Regarding the size of the Irish diaspora, we can see that the effects observed on the extensive margin for climate and health-related topics still exist in the intensive margin. However, while the share of Irish migrants does not influence the probability that a letter mentions economics (a feature shared by most letters, as shown in figure 3.5.1), it makes this topic much less salient. On average, increasing the share of Irish-born by 10 percentage points is associated with 3.5 fewer economic words (relative to a mean of 11.10). The composition of the county in terms of origins plays now an important role, as a higher share of foreign-born is associated with a higher importance of economics[48]. A plausible interpretation for this result is that places with the highest number of migrants from all origins are places of arrival,

[48] This pattern is also visible when we condition on the length of the letter, as displayed in table 3.A.17 in the Appendix.

Table 3.5.4: Migrant socio-economic environment and topic intensity (nb. of keywords)

	(1) Climate	(2) Health	(3) Economic	(4) Religion	(5) Politics	(6) Migration
Urbanization (%)	-0.469	-0.208	-0.485	-0.026	1.752***	0.403***
	(0.289)	(0.292)	(1.009)	(0.224)	(0.482)	(0.133)
Foreign-born (%)	3.270***	1.748**	10.095***	0.764	-0.558	-0.030
	(0.927)	(0.807)	(3.221)	(0.835)	(1.202)	(0.362)
Irish-born (%)	-9.692***	-7.595***	-35.365***	-3.926	-0.310	0.399
	(1.844)	(1.818)	(7.472)	(2.735)	(3.854)	(0.973)
Avg. income (std.)	-0.082	-0.000	-1.334***	-0.315***	-0.647***	-0.150***
	(0.126)	(0.111)	(0.382)	(0.110)	(0.145)	(0.052)
Income gap native-Irish (std.)	0.251*	0.165	0.831	0.358**	-0.521***	0.059
	(0.129)	(0.150)	(0.528)	(0.143)	(0.201)	(0.055)
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	2.66	2.74	11.10	2.35	3.29	1.06
R^2	0.05	0.06	0.10	0.06	0.10	0.05
Observations	2,659	2,659	2,659	2,659	2,659	2,659

Notes: The sample comprises all the letters sent from North America to Ireland during 1840-1929 with a precised location. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

where the economic activity is important and therefore where Irish migrants (and others) communicate mostly about whether work is hard to find or well-paid. This would reflect their first preoccupations when they arrive in the country.

Overall, this section shows that the socio-demographic environment surrounding migrants heavily influences the content of their letters. In particular, the size of the Irish diaspora and the economic conditions migrants face play an important role. We will extend this analysis looking at finer measures of migrants' integration, such as inter-marriages and naturalization rates.

3.6 Conclusion

Migrant letters are a unique source to shed light on migrants' lives, their hopes and struggles, and the diversity of feelings, from uprootedness to excitement about a new country and culture. By assessing migrants' communications first-hand, one can begin to understand when and how they embrace or reject natives' values, and how they transfer novel ideas to their communities of origin.

This paper opens the black-box of social remittances focusing on the Age of Mass Migration, one of the largest migration episodes in history. From 1840 to 1920, over 40 million Europeans emigrated to the Americas, drastically transforming both continents. We gather thousands of letters sent by Irish migrants living in North America to their family and friends left behind, which we combine with census data and information from passenger lists to create a novel dataset.

We exploit some of the most advanced artificial intelligence technologies combined with large language models to extract information from the letters and classify them into broad topics and themes. In particular, we focus on several dimensions including religious and political issues, as well as economic information among others. We assess the performance of the artificial intelligence and find that it produces a extremely similar classification to the one made by humans.

First, we conduct some descriptive analysis on the content of letters and provide some novel evidence about them. We find that economics and relationships are by far the most important topics, both in terms of frequency in the letters and intensity (i.e., the number of words). These findings are consistent with the fragile economic situation faced by most migrants as well as the need to preserve their ties with their family left behind. Migrants also discuss about politics and religion, two topics that were very important at the time given the cultural clash for Catholic migrants arriving in largely Protestant North America.

Second, we show that migrants' characteristics sharply influence the content of letters. In particular, the writer's gender plays a major role in the presence and intensity of several topics, with women being much more likely to talk about health and religion and less likely to discuss economic or political issues. The religious affiliation of the writer is also relevant. In particular, we find that Catholics discuss much more frequently religious ideas in their letters. Finally, the content of letters progressively becomes scarcer as the migrants spend more and more years abroad.

Third, beyond individual characteristics, we find that the local environment in which migrants live also heavily influences the content of the letters. For instance, the size of the local Irish diaspora is negatively associated with social remittances and the intensity of communications. These results suggest that migrants may substitute their connections back home with their local network, and contrary to what may be intuitive, social remittances may decrease after emigration surpasses a given threshold.

While our findings provide a preliminary assessment of migrants' communications and social remittances, we plan on expanding our investigation in multiple directions. First, we will construct more nuanced measures of social integration to understand whether they may undermine or foster the transfer of social remittances. Second, we will shed more light on the role of religious cleavages in US, in shaping the content of religious remittances. By identifying the likely affiliation of Irish migrants based on their names, we can characterize whether they reside in a religiously polarized community. Finally, we will exploit the staggered expansion of the Catholic church in the US as an exogenous shock to religiosity, to examine how these institutions may have foster religious remittances, and in turn, spark a revival in religiosity in migrants' communities of origin back in Ireland.

Appendix

3.A Additional tables

Table 3.A.1: Largest collections of letters.

Collection	N
J. A. Smyth	840
Moore Letters	234
Hurley Brothers Letters	128
Lough Sisters Letters	109
Flanagan Letters	100
B. O'Reilly	85
Maurice and Batt Wolfe Letters	71
Wright Letters	68
Kennedy and Walker Letters	50
Prendergast Letters	49
Doogan/Duggan Letters	48
Lalor Letters	47
Quinn "O'Brien" Letters	46
McCullogh-Hutchison Letters	42
Smith/ Smyth Letters	42
Williamson Letters	39
Griffin-Potter Letters	37
Hanlon/ O'Hanlon Letters	37
Malone-McHenry Letters	37
Kells Letters	33
McGuinness-Elliott Letters	33
Porter Letters	32
Teresa Lawlor Letters	31
Father Patrick Brosnan Letters	30
Patrick J. Monks Letters	30
Sinclair-Orr Letters	30
Sproule Letters	30
Other	
Collections with less than 30 letters	2066
No specific collection	2783
Unknown	155

Table 3.A.2: Distribution of sender locations at the country, state, and county level.

Sender country	N	Share (%)
United States	**3359**	**61.8**
Known state	3317	61.1
Known county	3264	60.1
Unknown county	53	1.0
Unknown state	42	0.8
Canada	**898**	**16.5**
Known state	866	15.9
Known county	813	15.0
Unknown county	53	1.0
Unknown state	32	0.6
Other countries	**977**	**18.0**
Unknown	**197**	**3.6**
Total	**5431**	**100.0**

All letters, 1840-1930

Table 3.A.3: Distribution of receiver locations at the country, state, and county level.

Receiver country	N	Share (%)
Ireland	**3138**	**57.8**
Known county	2901	53.4
Given	2716	50.0
Inferred	185	3.4
Unknown county	237	4.4
United States	**740**	**13.6**
Canada	**421**	**7.8**
United Kingdom	**71**	**1.3**
Unknown	**1061**	**19.5**

All letters, 1840-1930. 'Given' means that the county was mentioned in the letter or could be determined based on the address in the letter. 'Inferred' means that the county was determined using one of the inference method detailed in table 3.C.1.

Table 3.A.4: Missing information and migrant socio-economic environment

	(1) Sender's name	(2) Receiver's name	(3) Receiver's location	(4) Sender's name	(5) Receiver's name	(6) Receiver's location
Urbanization (%)	0.015	0.018	0.091***	0.010	0.018	0.091***
	(0.016)	(0.028)	(0.031)	(0.015)	(0.028)	(0.031)
Foreign-born (%)	0.023	0.014	-0.047	0.022	0.014	-0.047
	(0.039)	(0.085)	(0.094)	(0.039)	(0.085)	(0.094)
Irish-born (%)	-0.093	0.310	-0.238	-0.082	0.310	-0.238
	(0.091)	(0.219)	(0.212)	(0.090)	(0.219)	(0.212)
Average income (std.)	-0.007	-0.031***	-0.037***	-0.006	-0.031***	-0.037***
	(0.005)	(0.011)	(0.012)	(0.005)	(0.011)	(0.012)
Income gap native-Irish	-0.005	0.055***	0.005	-0.005	0.055***	0.005
	(0.006)	(0.011)	(0.015)	(0.006)	(0.011)	(0.015)
Decade FE	No	No	No	Yes	Yes	Yes
Country FE	No	No	No	Yes	Yes	Yes
Mean dep. variable	0.03	0.18	0.23	0.03	0.18	0.23
R^2	0.00	0.04	0.06	0.00	0.04	0.06
Observations	2,659	2,659	2,659	2,659	2,659	2,659

Notes: The sample comprises all the letters sent from North America to Ireland during 1840-1929 with a precised location. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.5: Balance table between all passengers and linked letter writers from IED only.

	All (N=1831746)		Linked (N=137)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
Age (at arrival)	26.33	9.38	24.16	9.67	-2.17	0.010
Female	0.45	0.50	0.26	0.44	-0.19	<0.001
Arrival in New York	0.95	0.22	0.96	0.19	0.02	0.335
Occupational income score (OIS)	10.40	5.18	13.01	8.38	2.61	0.006
Socio-eco. status index (SEI)	12.64	9.43	13.32	13.53	0.68	0.651
Earnings index score (ERS)	9.26	15.06	17.10	24.97	7.84	0.006
Educational index score (EDS)	4.19	5.60	4.95	8.03	0.76	0.396
Catholic name	0.22	0.41	0.05	0.22	-0.16	<0.001
Protestant name	0.05	0.22	0.26	0.44	0.21	<0.001
Neutral/Unknown name	0.73	0.44	0.69	0.47	-0.05	0.241

This table uses only letter writers from IED. Female is a dummy variable indicating whether the individual is a woman. Arrival in New York is a dummy indicating whether the individual's port of arrival is New York. OIS, SEI, ERS, and EDS are scores associated with occupation names as listed in the HISCO database. For religiosity, we use our measure based on name frequency in Irish District Electoral Divisions.

Table 3.A.6: Balance table between all passengers and linked letter writers from IMIRCE only.

	All (N=1831746)		Linked (N=120)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
Age (at arrival)	26.33	9.38	25.33	10.16	-1.00	0.287
Female	0.45	0.50	0.28	0.45	-0.17	<0.001
Arrival in New York	0.95	0.22	0.97	0.18	0.02	0.258
Occupational income score (OIS)	10.40	5.18	11.50	6.26	1.10	0.132
Socio-eco. status index (SEI)	12.64	9.43	11.65	10.83	-0.99	0.433
Earnings index score (ERS)	9.26	15.06	11.80	18.65	2.54	0.242
Educational index score (EDS)	4.19	5.60	4.04	3.61	-0.15	0.723
Catholic name	0.22	0.41	0.14	0.35	-0.07	0.023
Protestant name	0.05	0.22	0.13	0.34	0.08	0.010
Neutral/Unknown name	0.73	0.44	0.72	0.45	-0.01	0.845

This table uses only letter writers from IMIRCE. Female is a dummy variable indicating whether the individual is a woman. Arrival in New York is a dummy indicating whether the individual's port of arrival is New York. OIS, SEI, ERS, and EDS are scores associated with occupation names as listed in the HISCO database. For religiosity, we use our measure based on name frequency in Irish District Electoral Divisions.

Table 3.A.7: Migrant socio-demographics and topic presence (probit model)

	(1) Climate	(2) Health	(3) Economic	(4) Religion	(5) Politics	(6) Migration
main						
Female	0.352**	0.174	-0.072	0.424**	-0.381**	-0.065
	(0.170)	(0.189)	(0.257)	(0.186)	(0.163)	(0.162)
Age	-0.015**	0.009	0.007	-0.003	0.008	-0.015**
	(0.006)	(0.007)	(0.012)	(0.007)	(0.007)	(0.006)
Catholic name	-0.169	-0.315	-0.226	0.532*	0.209	-0.259
	(0.255)	(0.266)	(0.328)	(0.322)	(0.259)	(0.278)
Protestant name	0.604***	0.063	0.299	-0.054	0.152	0.183
	(0.223)	(0.244)	(0.322)	(0.192)	(0.212)	(0.207)
Years since arrival	0.001	-0.024**	-0.044***	-0.019*	-0.031***	-0.017
	(0.011)	(0.012)	(0.016)	(0.011)	(0.011)	(0.010)
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.64	0.78	0.93	0.68	0.62	0.47
R^2						
Observations	473	473	441	461	473	473

Notes: The sample comprises all the letters sent from North America to Ireland during 1840-1929 with a linked letter writer. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.8: Migrant socio-demographics and topic intensity (conditional on lenght)

	(1) Climate	(2) Health	(3) Economic	(4) Religion	(5) Politics	(6) Migration
Female	0.106	0.791**	-3.065***	0.233	-1.266**	-0.302**
	(0.319)	(0.345)	(0.806)	(0.393)	(0.546)	(0.129)
Age	-0.026*	0.006	-0.009	0.031	0.039	-0.015**
	(0.013)	(0.013)	(0.037)	(0.022)	(0.026)	(0.006)
Catholic name	-0.682	0.200	1.533	0.089	2.145	-0.493**
	(0.544)	(0.538)	(1.374)	(0.550)	(2.050)	(0.211)
Protestant name	1.203**	0.161	2.072	1.419*	0.400	-0.398**
	(0.569)	(0.348)	(1.488)	(0.727)	(1.063)	(0.181)
Years since arrival	0.054**	0.004	-0.007	-0.046*	0.021	0.001
	(0.022)	(0.019)	(0.058)	(0.025)	(0.042)	(0.009)
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	2.73	2.72	10.43	2.79	3.71	0.98
R^2	0.28	0.21	0.49	0.33	0.65	0.27
Observations	473	473	473	473	473	473

Notes: The sample comprises all the letters sent from North America to Ireland during 1840-1929 with a linked letter writer. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.9: Relationship between migrant socio-demographics and topic economic

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age	-0.001					0.001	0.001	0.001
	(0.001)					(0.001)	(0.001)	(0.001)
Female		-0.001				-0.003	-0.007	-0.008
		(0.028)				(0.028)	(0.031)	(0.031)
Catholic name			-0.047			-0.022	-0.026	-0.027
			(0.060)			(0.059)	(0.057)	(0.057)
Protestant name				0.036		0.041*	0.043*	0.030
				(0.025)		(0.025)	(0.026)	(0.033)
Years since arrival					-0.005**	-0.005**	-0.006**	-0.006**
					(0.002)	(0.002)	(0.002)	(0.002)
Decade FE	No	No	No	No	No	No	Yes	Yes
Country FE	No	No	No	No	No	No	No	Yes
R^2	0.00	0.00	0.00	0.00	0.02	0.02	0.03	0.03
Observations	480	473	480	480	480	473	473	473

Notes: The sample comprises all the letters sent from North America during 1840-1929 with a linked letter writer. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.10: Migrant socio-demographics and topic economic intensity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age	-0.152***					-0.060	-0.030	-0.028
	(0.045)					(0.047)	(0.045)	(0.045)
Female		-4.922***				-5.008***	-4.368***	-4.389***
		(0.888)				(0.987)	(1.088)	(1.089)
Catholic name			-1.219			0.494	1.881	1.840
			(1.823)			(1.766)	(1.686)	(1.694)
Protestant name				4.228**		3.494**	3.756**	3.347*
				(1.639)		(1.677)	(1.694)	(1.851)
Years since arrival					-0.304***	-0.295***	-0.173**	-0.176**
					(0.071)	(0.074)	(0.068)	(0.069)
Decade FE	No	No	No	No	No	No	Yes	Yes
Country FE	No	No	No	No	No	No	No	Yes
R^2	0.02	0.03	0.00	0.02	0.03	0.08	0.14	0.14
Observations	480	473	480	480	480	473	473	473

Notes: The sample comprises all the letters sent from North America during 1840-1929 with a linked letter writer. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.11: Migrant socio-economic environment and topic economic presence

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Urbanization (%)	-0.014						-0.014	0.009
	(0.010)						(0.015)	(0.018)
Foreign-born (%)		-0.063**					-0.077	-0.072
		(0.030)					(0.056)	(0.057)
Irish-born (%)			-0.017				-0.090	-0.143
			(0.068)				(0.125)	(0.130)
Avg. income score				0.000				
				(0.002)				
Income gap nat-irish					-0.001			
					(0.001)			
Female labor force (%)						-0.051		
						(0.033)		
Avg. income (std.)							0.014**	0.011*
							(0.007)	(0.007)
Income gap native-Irish (std.)							0.002	0.001
							(0.008)	(0.008)
Decade FE	No	No	No	No	No	No	Yes	Yes
Country FE	No	No	No	No	No	No	No	Yes
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
Observations	2,848	2,848	2,848	2,791	2,659	2,848	2,659	2,659

Notes: The sample comprises all the letters sent from North America during 1840-1929 with a precised location.
Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.12: Migrant socio-economic environment and topic economic intensity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Urbanization (%)	-3.929***						-0.773	-0.485
	(0.550)						(0.766)	(1.009)
Foreign-born (%)		-2.689					10.027***	10.095***
		(2.112)					(3.216)	(3.221)
Irish-born (%)			-8.286**				-34.718***	-35.365***
			(3.766)				(7.440)	(7.472)
Avg. income score				-0.623***				
				(0.096)				
Income gap nat-irish					-0.003			
					(0.072)			
Female labor force (%)						-22.235***		
						(2.360)		
Avg. income (std.)							-1.297***	-1.334***
							(0.374)	(0.382)
Income gap native-Irish (std.)							0.843	0.831
							(0.525)	(0.528)
Decade FE	No	No	No	No	No	No	Yes	Yes
Country FE	No	No	No	No	No	No	No	Yes
R^2	0.02	0.00	0.00	0.02	0.00	0.04	0.10	0.10
Observations	2,848	2,848	2,848	2,791	2,659	2,848	2,659	2,659

Notes: The sample comprises all the letters sent from North America during 1840-1929 with a precised location.
Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.13: Relationship between migrant socio-demographics and topic religion

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age	-0.004**					-0.000	-0.001	-0.001
	(0.002)					(0.002)	(0.002)	(0.002)
Female		0.141***				0.124**	0.140**	0.142**
		(0.048)				(0.050)	(0.055)	(0.055)
Catholic name			0.069			0.119	0.162*	0.167*
			(0.085)			(0.088)	(0.087)	(0.088)
Protestant name				-0.052		-0.010	-0.015	0.040
				(0.064)		(0.068)	(0.070)	(0.081)
Years since arrival					-0.010***	-0.010***	-0.007*	-0.007*
					(0.003)	(0.004)	(0.004)	(0.004)
Decade FE	No	No	No	No	No	No	Yes	Yes
Country FE	No	No	No	No	No	No	No	Yes
R^2	0.01	0.01	0.00	0.00	0.02	0.04	0.09	0.10
Observations	480	473	480	480	480	473	473	473

Notes: The sample comprises all the letters sent from North America during 1840-1929 with a linked letter writer. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.14: Migrant socio-demographics and topic religion intensity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age	-0.015					0.028	0.031	0.025
	(0.022)					(0.028)	(0.029)	(0.028)
Female		-0.055				-0.104	-0.192	-0.146
		(0.369)				(0.388)	(0.420)	(0.422)
Catholic name			-0.470			-0.010	0.085	0.177
			(0.503)			(0.478)	(0.485)	(0.492)
Protestant name				0.627		0.760	0.878	1.784*
				(0.730)		(0.765)	(0.821)	(1.043)
Years since arrival					-0.089***	-0.115***	-0.100***	-0.094***
					(0.020)	(0.030)	(0.033)	(0.032)
Decade FE	No	No	No	No	No	No	Yes	Yes
Country FE	No	No	No	No	No	No	No	Yes
R^2	0.00	0.00	0.00	0.00	0.02	0.03	0.07	0.08
Observations	480	473	480	480	480	473	473	473

Notes: The sample comprises all the letters sent from North America during 1840-1929 with a linked letter writer. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.15: Migrant socio-economic environment and topic religion presence

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Urbanization (%)	-0.069***						-0.105***	-0.144***
	(0.022)						(0.035)	(0.041)
Foreign-born (%)		-0.047					-0.095	-0.105
		(0.068)					(0.109)	(0.109)
Irish-born (%)			0.175				-0.490*	-0.402
			(0.141)				(0.250)	(0.255)
Avg. income score				-0.007**				
				(0.003)				
Income gap nat-irish					0.006**			
					(0.003)			
Female labor force (%)						-0.569***		
						(0.078)		
Avg. income (std.)							0.018	0.023
							(0.014)	(0.015)
Income gap native-Irish (std.)							0.077***	0.078***
							(0.018)	(0.018)
Decade FE	No	No	No	No	No	No	Yes	Yes
Country FE	No	No	No	No	No	No	No	Yes
R^2	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.05
Observations	2,848	2,848	2,848	2,791	2,659	2,848	2,659	2,659

Notes: The sample comprises all the letters sent from North America during 1840-1929 with a precised location.
Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.16: Migrant socio-economic environment and topic religion intensity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Urbanization (%)	-0.687***						-0.026	-0.392
	(0.176)						(0.224)	(0.274)
Foreign-born (%)		-0.225					0.764	0.678
		(0.685)					(0.835)	(0.834)
Irish-born (%)			0.307				-3.926	-3.099
			(1.419)				(2.735)	(2.759)
Avg. income score				-0.133***				
				(0.035)				
Income gap nat-irish					0.037*			
					(0.021)			
Female labor force (%)						-6.220***		
						(0.743)		
Avg. income (std.)							-0.315***	-0.268**
							(0.110)	(0.107)
Income gap native-Irish (std.)							0.358**	0.373**
							(0.143)	(0.145)
Decade FE	No	No	No	No	No	No	Yes	Yes
Country FE	No	No	No	No	No	No	No	Yes
R^2	0.00	0.00	0.00	0.01	0.00	0.03	0.06	0.06
Observations	2,848	2,848	2,848	2,791	2,659	2,848	2,659	2,659

Notes: The sample comprises all the letters sent from North America during 1840-1929 with a precised location.
Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.17: Migrant socio-economic environment and topic intensity (conditional on lenght)

	(1) Climate	(2) Health	(3) Economic	(4) Religion	(5) Politics	(6) Migration
Urbanization (%)	-0.723***	-0.435*	-1.818**	-0.688***	1.198***	0.256**
	(0.253)	(0.263)	(0.753)	(0.244)	(0.405)	(0.113)
Foreign-born (%)	2.764***	1.296*	7.439***	0.089	-1.661	-0.323
	(0.811)	(0.725)	(2.350)	(0.734)	(1.026)	(0.330)
Irish-born (%)	-7.272***	-5.437***	-22.681***	-0.286	4.959	1.799**
	(1.636)	(1.651)	(5.528)	(2.372)	(3.454)	(0.890)
Avg. income (std.)	0.103	0.164	-0.368	-0.054	-0.246*	-0.043
	(0.111)	(0.102)	(0.306)	(0.104)	(0.139)	(0.046)
Income gap native-Irish (std.)	0.153	0.078	0.318	0.259**	-0.734***	0.003
	(0.111)	(0.136)	(0.418)	(0.127)	(0.172)	(0.050)
Decade FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	2.66	2.74	11.10	2.35	3.29	1.06
R^2	0.21	0.20	0.46	0.24	0.37	0.26
Observations	2,659	2,659	2,659	2,659	2,659	2,659

Notes: The sample comprises all the letters sent from North America to Ireland during 1840-1929 with a precised location. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.B Additional figures

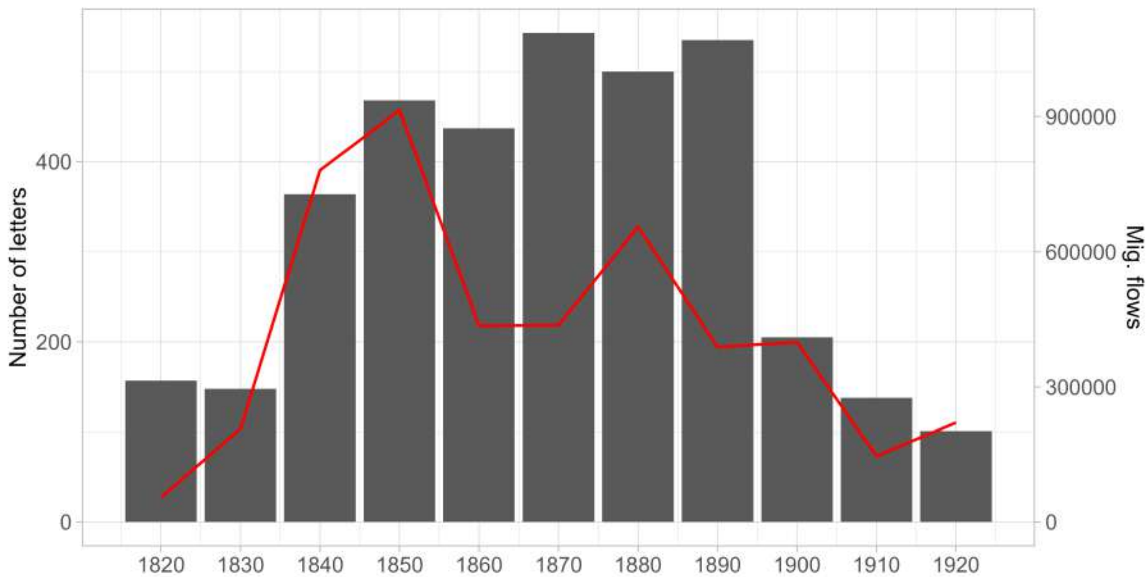


Figure 3.B.1: Number of letters in our sample and size of migration flows by decade. The left-axis corresponds to the number of letters sent from the United States (dark bars). This letters count excludes the J. A. Smyth collection, which is overrepresented in the 1880s and 1890s. The right-axis corresponds to the number of migrants from Ireland to the United States (red line).

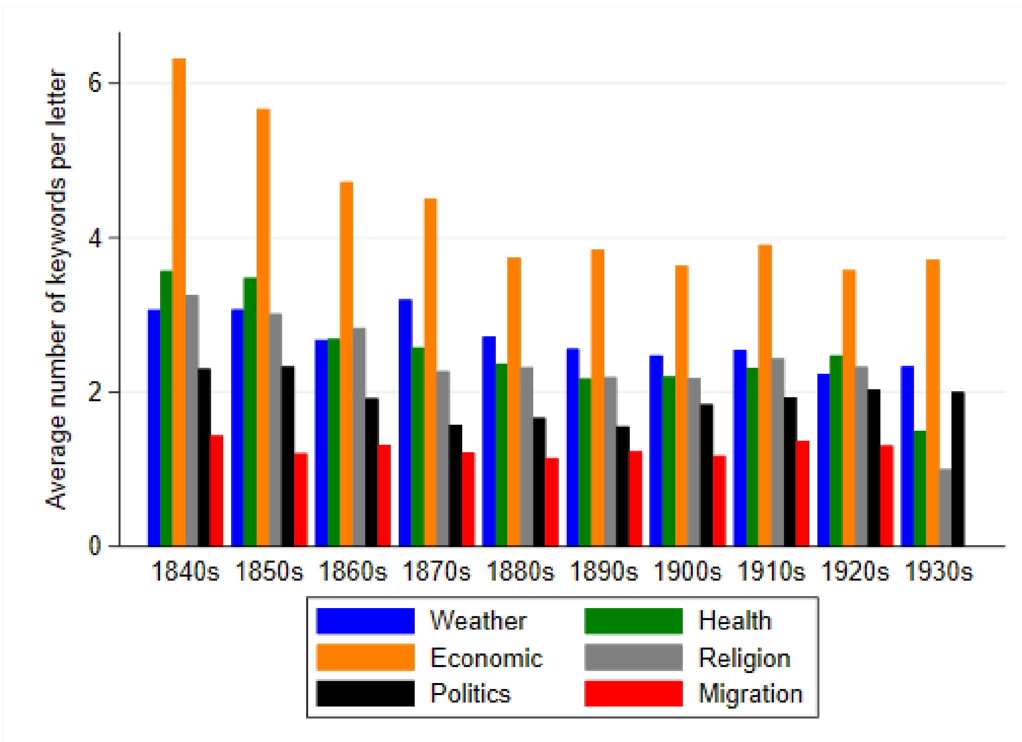


Figure 3.B.2: Theme frequency in the letters. This corresponds to the total number of keywords found in letters that have least one keyword in the corresponding theme, by decade.

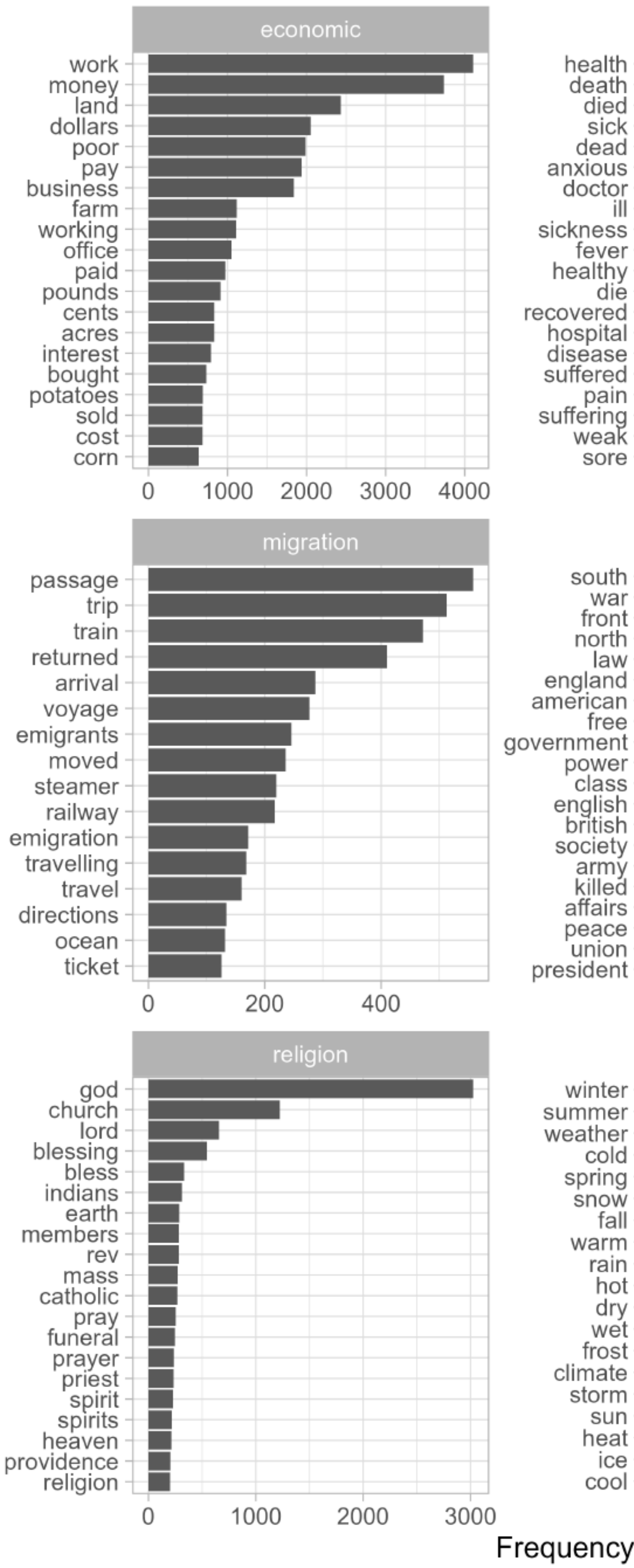


Figure 3.B.3: Keywords frequency in the letters.

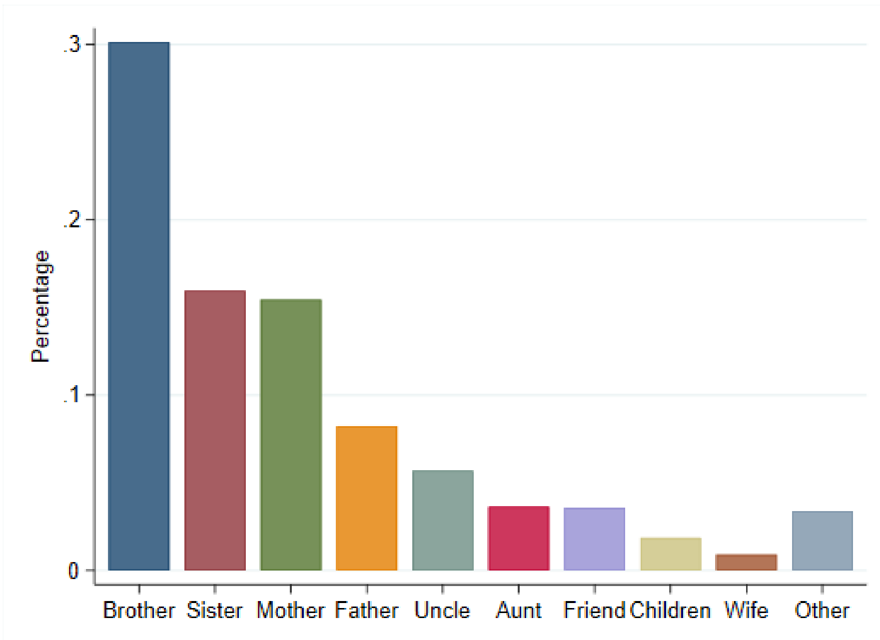


Figure 3.B.4: Distribution of the relationship of the letter receivers. This only considers letters in which this information is known.

3.C Geolocation of letters

3.C-I Harmonizing locations

The difficulty of this procedure varies depending on the source of the letters.

Letters provided by IMIRCE already have split the metadata into detailed categories, such as ”Names”, ”date”, and ”geographic information”. The latter contains the locations of the sender and the receiver together[49]. The only step we perform here is to split this unique variable into sender's and receiver's location.

Letters provided by IED are more challenging to process. The IED personnel that digitized and transcribed those letters added a one-sentence summary indicating (when available) the sender name, location, receiver name and location. However, the format of this sentence can importantly change between letters, making it hard to extract those components reliably in an automated way. Therefore, we proceeded in several steps.

First, we used ChatGPT 3.5 to split the one-sentence summary into sender name, sender location, receiver name, and receiver location. This output was checked and then post-processed to clean the text (e.g. exclude brackets, question marks, replace ”N.Orleans” by ”New Orleans”) and harmonize the output (e.g. replace ”Co.” and ”Co” by ”County”). The location is sometimes known but hard to geocode because it refers to a very small town. We use the website `https://www.townlands.ie/` to find this town and add more information.

Sometimes the sender location in the “source” variable is either absent or imprecise (e.g. only the country or the State) while we have some info in the letter header. When this is the case, we give this sample of data to ChatGPT with: letter ID, location of the receiver, first 200 characters in the letter (exact prompts are provided in section 3.G). We include the location of the receiver to be able to tell ChatGPT that it shouldn't take into account this location when looking for the location of the sender. We do the same thing with the last 200 characters of the letter.

3.C-II Missing information

Sometimes, the text of the letter doesn't contain any information on the receiver location, or only vague locations, such as ”Ireland”. To address this issue, we take advantage of the fact that all letters are not independent from each other but belong to collections. Therefore, we can use the location information of other letters from the same location to infer the location

[49] For example: ”Philadelphia (city),Philadelphia (county),Pennsylvania,United States,Moycraig Hamilton (townland),Antrim (county),Ireland”

when it is missing. We use four different strategies to infer the receiver's location at the county level. For the first 3, we use various information on the sender's name, county, and receiver's name to fill missing receiver's county. Table 3.C.1 details those strategies and shows some examples.

Strategy 1: Use sender's name and county (within 10 years), and receiver's name. Below, replace the missing value with "County Down".				
Sender	**Year**	**Sender county**	**Receiver**	**Receiver county**
James Ryan	1881	Allegheny	John Smith	County Down
James Ryan	1885	Allegheny	John Smith	??
James Ryan	1890	Allegheny	John Smith	County Down

Strategy 2: Use sender's name and collection (within 10 years). Below, replace the missing value with "County Down".				
Sender collection	**Year**	**Sender**	**Receiver**	**Receiver county**
J. Ryan collection	1881	James Ryan	Bryan May	County Down
J. Ryan collection	1885	James Ryan	John Smith	??
J. Ryan collection	1890	James Ryan	John Lennon	County Down

Strategy 3: Use sender's name and county (within 10 years). Below, replace the missing value with "County Down".				
Sender county	**Year**	**Sender**	**Receiver**	**Receiver county**
Allegheny	1881	James Ryan	Bryan May	County Down
Allegheny	1885	James Ryan	John Smith	??
Allegheny	1890	James Ryan	John Lennon	County Down

Table 3.C.1: Refined procedure to infer receiver's location if missing

3.C-III Geolocation

Finally, we use ArcGIS to geocode the addresses, so that we obtain latitude and longitude as well as county and State names. Geocoding addresses can be challenging, especially when names are vague or correspond to several places. Therefore, we also used Google's geocoding API and compared its results to ArcGIS. In the large majority of the cases, they return the same county and state information. In cases where their results differ, we manually check the original location in the letters and we mark those cases as uncertain. In the analysis where we use this location information, we run regressions with and without those uncertain locations.

3.D Linking writers to passenger lists

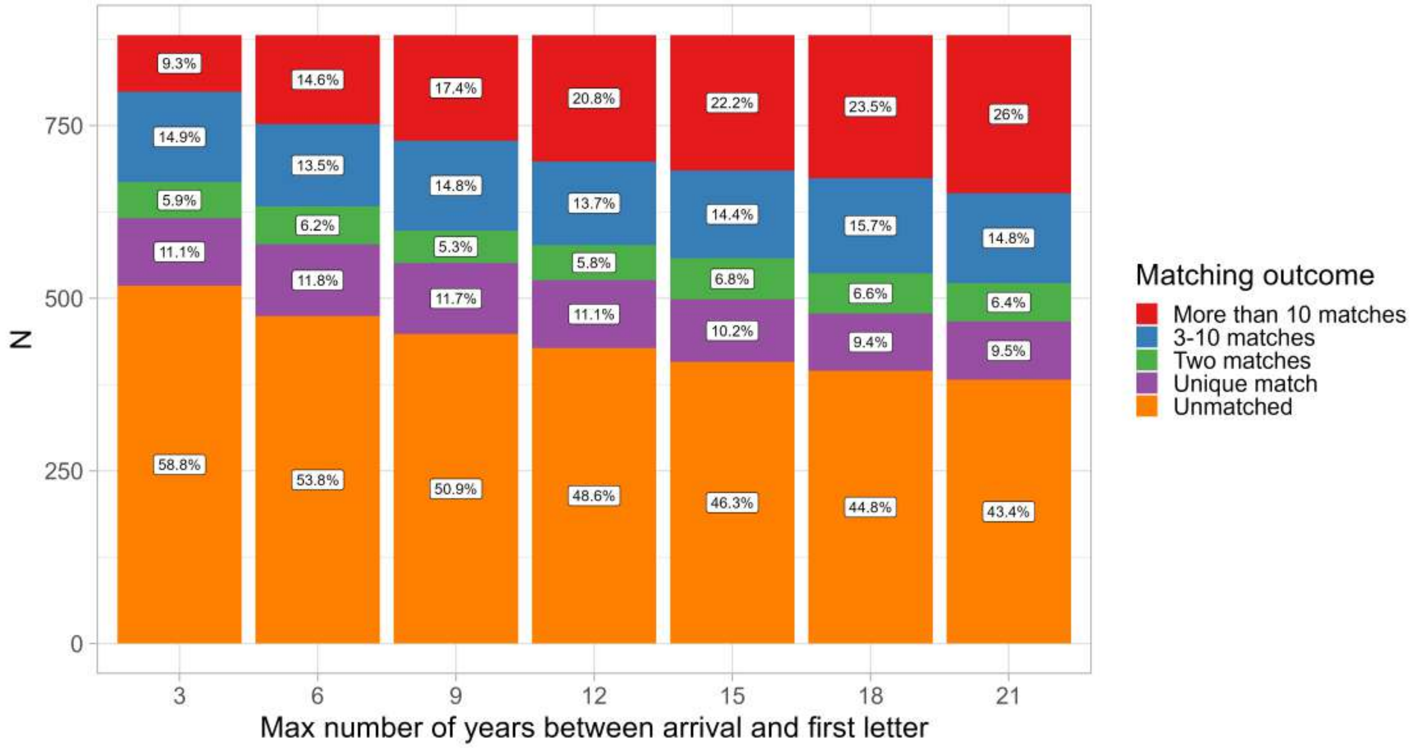


Figure 3.D.1: "Naive" linking: keep the same parameter for all letter writers. Each bar corresponds to a specific value for the maximum distance between year of arrival and year of letter. This is applied to the full sample that is already matched based on name similarity.

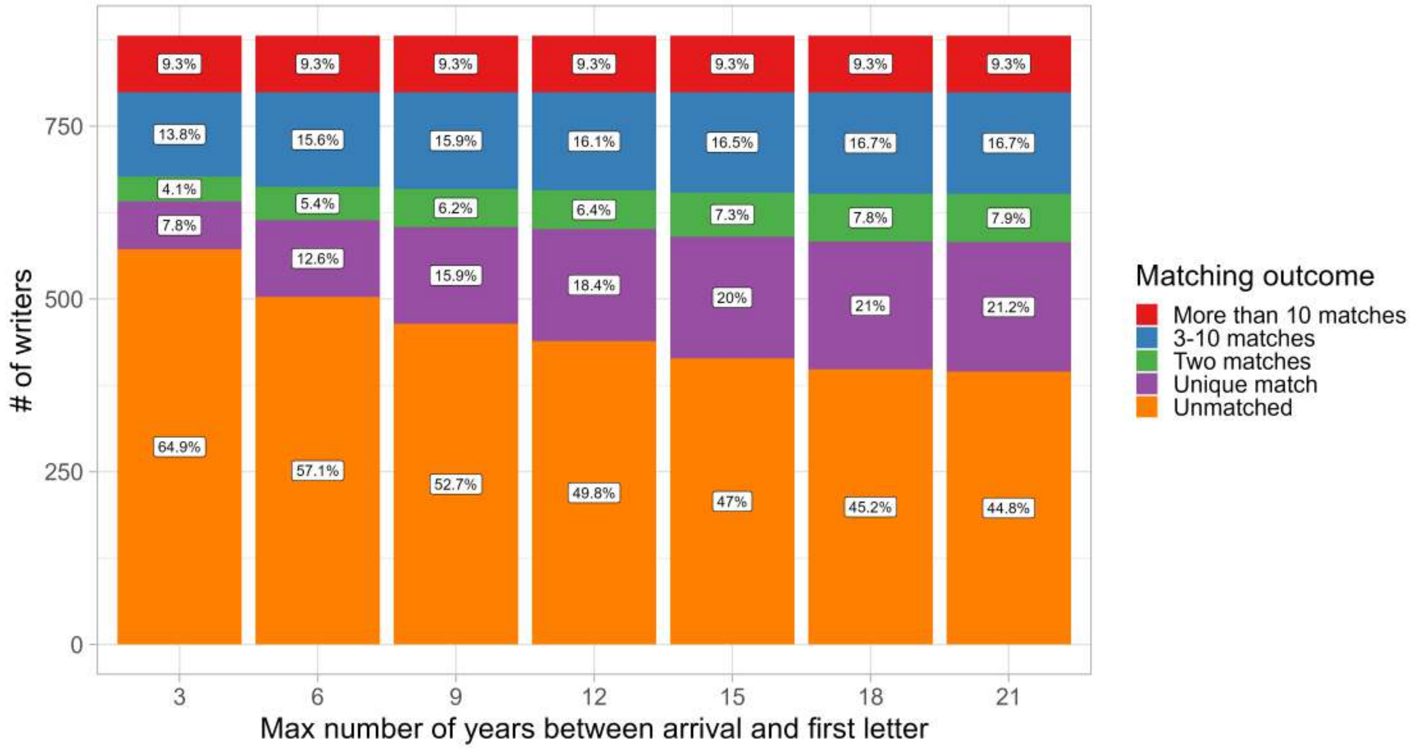


Figure 3.D.2: "Incremental" linking: if no link is found, relax the parameter on the maximum distance between year of arrival and year of letter.

3.E Content classification

Table 3.E.1: Main themes and sub-themes by topic

Topic	Theme	Sub-theme
Climate	Climate Conditions & Geography	Contrasting Climates & Geography: Ireland vs. North America
		Extreme Wheather Events (Harsh Winters, Snowstorms)
		Regular Weather Events
	Climate's Impact on Health and Well-being	Climate and Health Concerns
		Weather as a Shared Familial Experience
	Climate and Geography's Effect on Livelihoods	Impact of Climate and Geography on Agriculture
		Climate as a Factor in Migration and Settlement Choices
	Adaptation to New Climates	Adapting to New Climates & Geographies and Coping Strategies
		Clothing and Practical Advice for Coping with Weather
	Urban and Regional Climate Challenges	Urban Climate Challenges (Sanitation, Disease Spread)
		Diverse Regional Weather Conditions and Experiences
Health	Physical Health	Describing Ailments and Illnesses
		Expressing Improvement from Health Condition
	Mental and Emotional Health	Emotional and Spiritual Well-being in a Positive Manner
		Emotional and Spiritual Well-being in a Negative Manner
	Health Communication and Reassurance	Inquiring About the Health of Loved Ones
		Reassurance About Personal Health
		Expressing Concern for Family Health
	Coping with Death and Dying	Coping with Death and Dying of a Relative at Destination
		Coping with Death and Dying of a Relative at Origin
	Healthcare Practices and Access	Sharing Remedies and Treatments
		Healthcare Access and Limitations
		Advice on Avoiding Illnesses
	Factors Influencing Health	Impact of Hard Work on Health
		Climate and Its Impact on Health
		Health as a Motivation for Migration
Economics	Wages and Salaries	Specific occupations (e.g., laborers, mechanics)
		Regional wage variations
	Employment Opportunities and Scarcity	Availability of work by season or industry
		Unemployment and hardship during downturns
	Cost of Living	Price of food, housing, and essential goods
		Comparison between regions (Ireland vs. North America)
	Financial Support and Remittances	Sending money back to Ireland
		Requesting financial aid from relatives
	Working Conditions	Labor relations and strikes
		Employer-worker dynamics and exploitation
	Entrepreneurship and Social Mobility	Starting businesses or acquiring land
		Rising through job ranks
	Economic Hardship and Resilience	Struggles with unemployment, low wages, and rent
		Coping strategies during economic downturns
	Advising on Economic Migration	Financial advice on whether to migrate
		Warning against unrealistic expectations
	Economic Impact on Personal Lives	Decisions influenced by financial stability
		Marriage, education, and family, linked to income
Religion	Religious Practices in Daily Life	Church attendance
		Prayer meetings
		Bible readings and hymns
	Religious Education	Importance of Sunday schools
		Concerns over denominational influences (e.g., Methodists)
	Religious Diversity and Exposure to New Denominations	Encounter with Protestant sects (Quakers, Methodists)
		Firmness in Catholic identity despite exposure
	Religious Change and Conversion	Experiences of religious conversion (e.g., to Methodism)
		Changing denominational affiliations
	Religion as a Community Anchor	Religious services as a means of community building
		Support networks centered around churches
	Expressions of Personal Faith	Reflections on personal salvation
		Appeals for spiritual support (prayers from family)
	Second Great Awakening and Revivals	Exhortations and sermons
		Reactions ranging from engagement to skepticism
	Anti-Catholic Sentiment	Facing prejudice and discrimination in a Protestant society
		Criticism of religious intolerance
	Religion as a Source of Comfort	Finding solace in faith during difficult times
		Prayers for divine guidance
	Religion and Identity	The intertwining of Catholicism and Irish identity
	Religious Hypocrisy	Criticism of those whose actions contradict religious beliefs

Table 3.E.2: Main themes and sub-themes by topic (cont.)

Topic	Theme	Sub-theme
Politics	American Politics.	Presidential elections Political party loyalty (especially the Democratic Party) Influence of political events on daily life
	Irish Politics and Nationalism	Irish nationalist movements (Repeal of the Act of Union) Influence of figures like Daniel O'Connell Continued interest in political developments in Ireland
	Slavery and Abolition	Support for abolitionism Anti-slavery stance during the American Civil War Divisions within the Irish-American community regarding slavery
	Anti-Irish Sentiment and Nativism	Challenges due to nativist movements Anti-Irish Sentiment and Nativism. Prejudice against Irish immigrants
	Social Reforms	Temperance movements (anti-alcohol) Criticism of economic and social conditions affecting migrants
	Gender and Social Class	Class mobility in the U.S. Roles and expectations for women in the migrant context
	International Relations and Conflicts	Concern over U.S. relations with Britain Reflections on conflicts like the Crimean War
	Transatlantic Activism.	Support for Irish causes from abroad Irish migrants' involvement in political activism to aid their homeland
Migration	Push and Pull Factors for Migration	Economic hardships in Ireland Promise of better opportunities in America (land, wages, etc.) Escaping political or religious oppression
	Chain Migration	Following family or friends who had already migrated Establishing networks for newcomers
	Family Separation and Reunification	Pain of leaving loved ones behind Efforts to bring family members over
	Journey Experiences	Difficult sea voyages Costs and logistics of migrating
	Adapting to a New Life	Cultural adjustment to life in America Learning new customs and language
	Maintaining Connections with Ireland	Sending letters, remittances, and updates Keeping Irish traditions alive abroad
	Social Networks and Support	Importance of Irish communities and networks Role of churches and social organizations
	Employment and Economic Challenges	Opportunities in skilled and unskilled labor Economic challenges and competition for jobs
	Homesickness and Longing for Return	Emotional toll of migration Dreams of returning to Ireland but facing realities of settling permanently
	Disillusionment and Reality	Unrealistic expectations vs. harsh realities in America Some returning to Ireland after failed attempts
	Advice and Recommendations for Future Migrants	Encourage migration. Practical tips on traveling and settling Discourage migration. Warnings about challenges and hardships
	Role of Weather and Climate	Climate as a factor in settling in different regions Adapting to new weather conditions in America
	Migration and Identity	Balancing Irish identity with Americanization The role of religion and culture in preserving Irishness in diaspora
	Impact of Migration on Family Dynamics	Changing gender roles due to migration Impact on family responsibilities and expectations
Relationships	Family Connections	Maintaining Family Connections Across Distances Emotional Impact of Separation Chain Migration and Family Reunification Plans for Family Reunification
	Financial and Material Support	Financial Support and Remittances Social Networks and Support Systems
	Family Dynamics and Structure	Changing Family Dynamics Importance of Marriage and Family Formation Intergenerational Relationships Generational Differences
	Community and Social Networks	The Importance of Community and Neighbors Gossip and Social Commentary
	Cultural Identity and Traditions	Maintaining Irish Identity and Traditions Courtship, Marriage, and Family Building Social Etiquette and Expectations

Below are the keywords considered in each topic:

- **climate** : climate, cloudy, cold, colder, cool, dry, dust, fall, flooded, frost, frostbitten, frosty, froze, frozen, heat, hot, ice, icebound, rain, raining, showers, snow, spring, squall, storm, storms, stormy, sultry, summer, summers, sun, sunshine, thermometer, thunder, thunderstorm, warm, warn, weather, wet, wettest, winter, winters
- **health** : anxious, bilious, blood, brain, cancer, colds, corpse, dead, death, debility, decease, die, died, disease, diseases, doctor, dying, dysentery, epidemic, erysipelas, fainted, fever, headache, health, healthy, hospital, ill, influenza, limbs, lungs, measles, medicine, mental, pain, painfull, pills, quarantine, quinine, recover, recovered, recovering, remedies, resane, rheumatism, scarlet, seasick, sick, sickly, sickness, soares, sore, starvation, stomach, stroke, suffered, suffering, sunstroke, throat, toothace, typhoid, typhus, unwell, veins, weak, wounded
- **economic** : acres, agriculture, apprenticeship, assets, attorney, auditors, avail, bales, bank, banking, bankruptcy, banks, barber, bargain, bill, bills, blacksmiths, bond, bonds, bought, brokers, business, businessmen, busness, bussiness, buy, calf, capital, carpenter, carpenters, carriage, cash, cashed, cattle, cent, cents, charge, charged, cheap, chicken, chickens, chief, clerk, client, coal, colonist, commodities, consumption, contractor, corn, cost, cotton, cow, cows, credit, crop, crops, customers, cutting, debt, debts, depression, discount, dividend, dollar, dollars, dressmakers, driving, dull, dwelling, earning, earns, economy, eggs, employ, employed, employment, engineer, especulative, estate, expenses, factory, famine, farm, farmers, farming, financial, firm, fisheries, fortune, fund, funds, gain, gold, goods, grain, grains, ground, heirs, hire, hogs, horse, horses, installment, insurance, interest, investment, invoices, job, keeper, labor, laborers, labour, laboured, lamb, land, lawsuit, lawyer, lawyers, legal, loaned, loss, lounging, machine, machines, manufacturer, manufactures, manufacturing, market, materials, mechanics, merchant, mines, mining, misfortune, money, notary, oats, office, officers, oxen, paid, partner, partners, partnership, pay, paying, payment, per cent, percent, percentage, pernsion, plough, ploughs, plow, plowing, poor, pork, porter, potatoes, pound, pounds, poverty, price, prices, produce, profis, profit, profits, properties, property, prospect, prosperity, prosperous, purchase, purchased, purchasing, qualifications, rate, receipt, recession, rent, rent, renting, repay, rich, salary, salarys, sales, sawmill, sell, selling, service, sewing, share, shares, sheep, shilling, shillings, shipped, shop, shovel, silver, societies, sold, sowing, spade, speculated, speculative, speculators, stagnation, sterling, stock, stocks, store, sum, surplus, tax, taxes, taylors, teacher, teaching, timber, trade, trademen, traders, trades, tradesmen, transaction, transactions, treasury, trustee, tunnels, unpaid, vacancy, vegetables, wage, wages, warn, wealth, wealthy, wheat, wholesale, work, worked, working, workmen

- **religion** : abyss, advocates, almighty, apostole, apostolic, belief, bishop, bless, blessed, blessing, blessings, burying, cathedral, catholic, catholics, christ, christian, christians, church, clergyman, communion, comunion, converted, converts, creed, darkness, denominations, devotion, disciples, divine, earth, eternal, eternity, faith, faithful, funeral, giver, god, gods, graces, graveyard, guard, guidance, heaven, heavers, immortality, indians, jesus, lord, mass, members, mercies, mercy, methodist, methodists, minister, missionary, mohommet, papist, pledge, praise, pray, prayer, praying, preached, preacher, preaching, presbyterian, presbyterians, priest, procession, protestant, protestants, providence, realm, religion, religious, resurrection, rev, revivals, scripture, sect, sermon, sinful, solemn, soul, spirit, spirits, spiritual, supper, temptation, theolog, worship
- **politics** : affairs, alliance, American, americans, anarchists, antislavery, army, assault, assembly, attacking, ballot, battle, biased, blacks, bloodshed, bloody, bondage, brave, britain, british, brotherhood, candidate, capitol, chamber, citizens, citizens, civil, class, club, colonel, colony, coloured, comission, command, confederate, Congress, conservative, constitution, convention, coronel, corruption, courts, crime, crown, darkness, declaration, defence, defenders, democracy, democrats, despotisms, duel, electing, election, elections, electoral, emancipation, enemies, enemy, england, english, envy, evil, execration, exile, exile, exiled, federalist, fenians, force, foreigners, forts, free, freedom, french, front, gaelic, government, gun, guns, harcourt, hospitality, Independence, independent, institution, institutions, irelanders, irish american, irish flag, irish republic, jail, killed, king, kingdoms, knownothing, law, laymen, liberty, lynch, military, murder, nation, national, native, navy, negro, North, Oligarchy, orange, orangemen, orangenism, Parliament, patriotic, patriots, peace, peacefully, persecuted, plantation, politics, popists, power, prejudice, president, prince, prisoners, proscription, protest, quarrelling, queen, race, rank, rape, rebel, rebellion, reform, representative, republic, republican, revolution, rights, riot, riotting, royal, royalty, rule, rulers, safety, secceed, senator, shabby, shot, slave, slavery, slaves, society, south, strike, sword, threaten, threatening, treason, treason, troops, troublesome, unanimous, union, united, uproar, vengeance, volunteered, vote, voter, voting, war, yankees, yankeys
- **migration** : accompany, arrival, arriving, come back, come home, come out here, come to Ireland, come to see you, coming here, coming home, coming to Europe, crossing, directions, emigrants, emigrate, emigration, encouragement, go out to, go to you, going back, going home, going out to america, left home, migrate, moved, not coming, ocean, parting to, passage, passages, railway, ready to come, return to, returned, route, stay home, steamer, steamers, take a trip, ticket, train, travel, travelling, trip, voyage, wagons, you to come

3.F Names and religion

The US and Canadian censuses and passenger lists do not record the religion of individuals. Therefore, we infer whether someone is Catholic, Protestant, or has unknown religiosity based on their name. To do so, we use data from the Irish full count census of 1901, which records first names and self-declared religion, among other things. We compute the most popular names among Catholics and Protestants separately, based on whether they live in a District Electoral Division (DED) that has a Catholic or Protestant majority. Therefore, we end up with four distinct lists of the 50 most popular names for men and women: Catholics in Catholic-dominated places (C-C)[50], Protestants in Catholic-dominated places (P-C), Catholics in Protestant-dominated places (C-P), and Protestants in Protestant-dominated places (P-P) (see tables 3.F.1 and 3.F.2 for the list of most popular names for women and men, respectively).

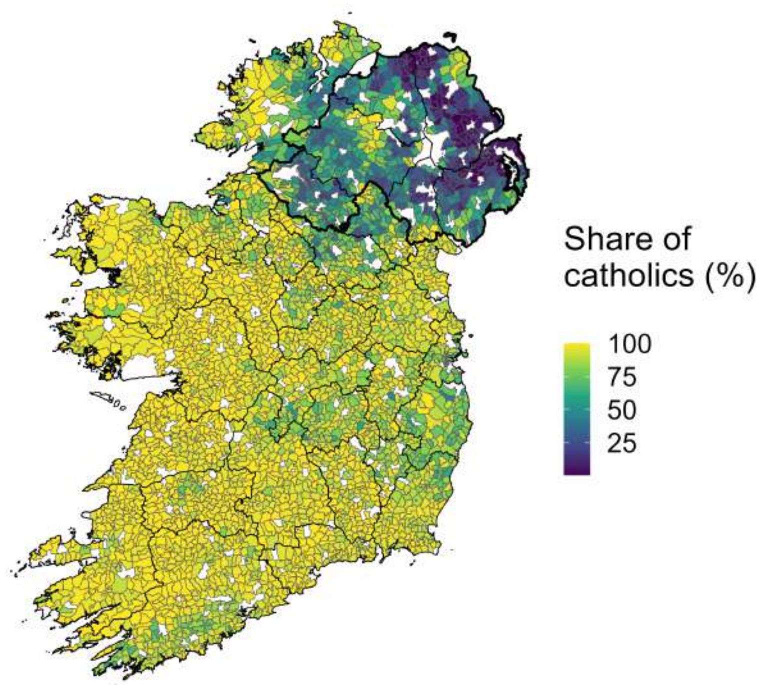


Figure 3.F.1: Distribution of Catholics in 1901
Note: The plot shows the share of catholics per District Electoral Division (DED). The black border distinguishes current Ireland and Northern Ireland. Data from the Irish census of 1901.

[50] A DED is considered "Catholic-dominated" ("Protestant-dominated") if more than 70% of its population self-declare "Catholics" ("Protestants").

From this data, we classified names as ”Catholic” if they appear in C-C or C-P and if their frequency is more than 5 times higher than their appearance in P-C or P-P. For instance, we find that ”Bridget” is the second most popular name for Catholics women living in Catholic places (11.5%), and the 7th most popular name for Catholics women living in Protestant places (3.7%). On the other hand, it doesn't appear at all in the most popular names for Protestant women, no matter their place of residence. Therefore, we classify ”Bridget” as a Catholic name. The reverse procedure is applied to classify a name as ”Protestant”. If a name does not appear in the list of most popular names or is popular both for Catholics and Protestants[51], then we classify its religiosity as ”unknown”.

Figures 3.F.2 and 3.F.3 show that the share of Catholics (Protestants) that we infer based on names is correlated with the actual share of Catholics (Protestants) that is computed based on the self-reported religion.

[51] For instance, ”Mary” is the most popular female name, no matter the place of residence and the self-declared religion.

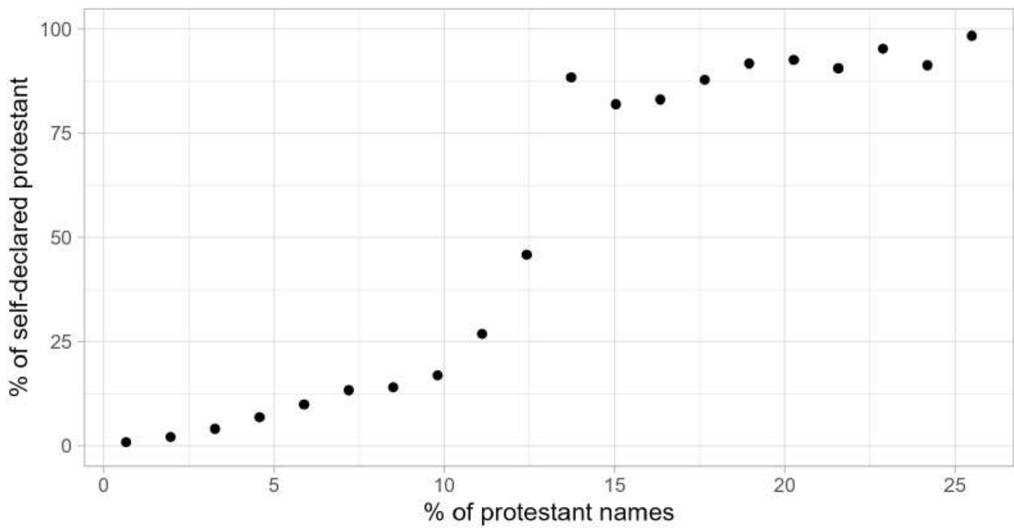


Figure 3.F.2: Correlation between the share of individuals who self-declared "Protestant" and the share of individuals who were inferred "Protestant" based on their names. Bin scatter plot using 20 bins. Authors' own calculation based on data at the District Electoral Division (DED) level from the 1901 Irish Census.

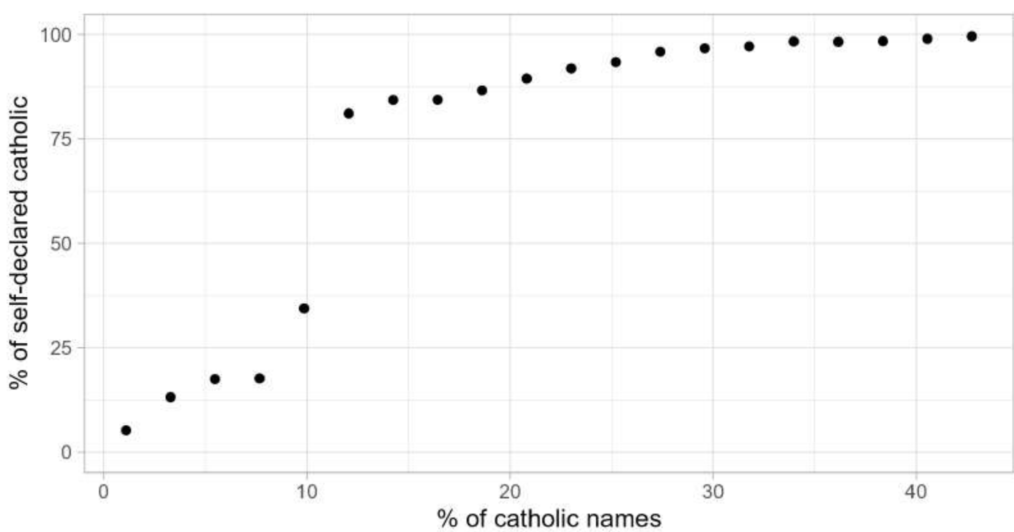


Figure 3.F.3: Correlation between the share of individuals who self-declared "Catholic" and the share of individuals who were inferred "Catholic" based on their names. Bin scatter plot using 20 bins. Authors' own calculation based on data at the District Electoral Division (DED) level from the 1901 Irish Census.

Table 3.F.1: Most popular female names by self-declared religion and majority religion of place of residence.

Rank	Catho. in Catho. place	Prot. in Catho. place	Catho. in Prot. place	Prot. in Prot. place
1	Mary (19.1%)	Mary (5.1%)	Mary (14.4%)	Mary (7.9%)
2	Bridget (9.9%)	Elizabeth (4.4%)	Margaret (4.9%)	Jane (5.9%)
3	Margaret (6.1%)	Jane (3.8%)	Sarah (4.8%)	Sarah (5.5%)
4	Ellen (5.7%)	Sarah (3.5%)	Annie (4.7%)	Elizabeth (4.3%)
5	Catherine (4.7%)	Margaret (3.4%)	Catherine (4.7%)	Annie (4.1%)
6	Kate (4.5%)	Annie (3.1%)	Ellen (4.2%)	Margaret (4.1%)
7	Anne (4%)	Ellen (1.9%)	Elizabeth (3.4%)	Maggie (3.7%)
8	Annie (2.6%)	Eliza (1.9%)	Bridget (3.3%)	Agnes (3.4%)
9	Maggie (2.2%)	Anne (1.6%)	Maggie (3.3%)	Lizzie (3.3%)
10	Julia (2%)	Lizzie (1.5%)	Jane (3.1%)	Martha (2.5%)
11	Elizabeth (1.6%)	Susan (1.4%)	Lizzie (2.6%)	Ellen (2.3%)
12	Sarah (1.5%)	Mary Jane (1.3%)	Rose (2.3%)	Eliza (2.1%)
13	Johanna (1.3%)	Isabella (1.3%)	Mary Ann (1.8%)	Isabella (2%)
14	Lizzie (1.2%)	Martha (1.3%)	Kate (1.4%)	Mary Jane (1.7%)
15	Eliza (1.2%)	Catherine (1.3%)	Agnes (1.3%)	Minnie (1.2%)
16	Mary Anne (1.2%)	Maggie (1.1%)	Alice (1.3%)	Mary Ann (1.2%)
17	Hannah (1.2%)	Emily (1.1%)	Eliza (1.3%)	Catherine (1%)
18	Katie (1.1%)	Mary Anne (1.1%)	Anne (1.3%)	Margret (0.8%)
19	Jane (1.1%)	Alice (1.1%)	Susan (1.3%)	Susan (0.8%)
20	Rose (1%)	Rebecca (0.9%)	Margret (0.9%)	Matilda (0.8%)
21	Norah (1%)	Maria (0.9%)	Isabella (0.9%)	Rachel (0.8%)
22	Alice (0.9%)	Frances (0.9%)	Mary Jane (0.9%)	Sarah Jane (0.7%)
23	Maria (0.9%)	Kate (0.8%)	Ann (0.9%)	Maria (0.6%)
24	Margret (0.9%)	Charlotte (0.7%)	Minnie (0.7%)	Eliza Jane (0.6%)
25	Nora (0.8%)	Florence (0.6%)	Mary Anne (0.7%)	Emily (0.6%)
26	Ellie (0.7%)	Agnes (0.6%)	Mary A (0.7%)	Alice (0.6%)
27	Susan (0.5%)	Hannah (0.6%)	Martha (0.7%)	Hannah (0.5%)
28	Mary A (0.5%)	Kathleen (0.6%)	Teresa (0.7%)	Rebecca (0.5%)
29	Teresa (0.4%)	Fanny (0.5%)	Cathrine (0.6%)	Anne (0.5%)
30	Agnes (0.4%)	Matilda (0.5%)	Kathleen (0.5%)	Anna (0.4%)

Statistics using the 1901 Irish Census at the district electoral division (DED) level. All individuals are considered. A DED is considered Catholic (Protestant) if at least 70% of its population declare themselves Catholics (Protestants).

Table 3.F.2: Most popular male names by self-declared religion and majority religion of place of residence.

Rank	Catho. in Catho. place	Prot. in Catho. place	Catho. in Prot. place	Prot. in Prot. place
1	John (16%)	William (10.2%)	John (15.2%)	John (11.8%)
2	Patrick (12%)	John (9.4%)	James (12%)	William (11.3%)
3	James (9.6%)	James (6.3%)	Patrick (8.9%)	James (11.2%)
4	Michael (9.2%)	Thomas (5.7%)	Thomas (5.6%)	Robert (7.4%)
5	Thomas (8%)	Robert (5.1%)	William (5.5%)	Thomas (5.8%)
6	William (4.5%)	George (4.4%)	Hugh (3.2%)	Samuel (5.2%)
7	Edward (2.1%)	Richard (2.3%)	Joseph (3.2%)	David (3.2%)
8	Daniel (2.1%)	Samuel (2.2%)	Edward (2.7%)	Joseph (2.7%)
9	Peter (2.1%)	Joseph (2.2%)	Daniel (2.7%)	George (2.6%)
10	Joseph (1.9%)	Henry (2.1%)	Michael (2.7%)	Alexander (2.5%)
11	Martin (1.8%)	Charles (1.8%)	Henry (2.4%)	Hugh (2.3%)
12	Denis (1.6%)	Edward (1.7%)	Charles (2.3%)	Henry (1.4%)
13	Timothy (1.3%)	David (1.2%)	Robert (2.1%)	Andrew (1.4%)
14	Richard (1.1%)	Alexander (1.2%)	Francis (2.1%)	William John (1%)
15	Francis (1.1%)	Arthur (0.8%)	Bernard (1.7%)	Charles (0.9%)
16	Pat (1%)	Francis (0.7%)	Peter (1.6%)	Edward (0.9%)
17	Jeremiah (0.9%)	Andrew (0.7%)	George (1%)	Richard (0.7%)
18	Bernard (0.9%)	Frederick (0.6%)	Arthur (1%)	Francis (0.7%)
19	Charles (0.8%)	Hugh (0.5%)	Samuel (0.9%)	Wm John (0.5%)
20	Hugh (0.7%)	Albert (0.5%)	Alexander (0.8%)	Matthew (0.5%)
21	Owen (0.6%)	Alfred (0.5%)	David (0.7%)	Daniel (0.5%)
22	Andrew (0.6%)	Benjamin (0.4%)	Richard (0.5%)	Arthur (0.4%)
23	Cornelius (0.6%)	Walter (0.4%)	Felix (0.5%)	William James (0.4%)
24	David (0.5%)	William John (0.3%)	Andrew (0.4%)	Albert (0.3%)
25	Henry (0.5%)	Frank (0.3%)	Martin (0.4%)	Isaac (0.3%)
26	Maurice (0.5%)	Michael (0.3%)	Owen (0.4%)	Willie (0.3%)
27	Edmond (0.5%)	Herbert (0.3%)	William John (0.3%)	Alfred (0.3%)
28	Christopher (0.5%)	Harry (0.3%)	Frank (0.3%)	Adam (0.3%)
29	George (0.5%)	Willie (0.3%)	Denis (0.3%)	Robert J (0.3%)
30	Robert (0.5%)	Peter (0.3%)	Matthew (0.3%)	Robert James (0.2%)

Statistics using the 1901 Irish Census at the district electoral division (DED) level. All individuals are considered. A DED is considered Catholic (Protestant) if at least 70% of its population declare themselves Catholics (Protestants).

3.G ChatGPT prompts used

We used ChatGPT 3.5 and 4.0 on the web interface. The following prompts were used to extract the location of senders and receivers from the text:

- **Refining broad location of sender using first lines of text**
  I have a list of letters sent by Irish migrants (mostly in the US and Canada, but not only) to their family/friends in Ireland (though not always). This list includes the letter link at the beginning, some broad information about the location of the letter writer (either the state or the country), and then the first words of the letter.
  Usually, those first words contain the location of the letter writer and of the letter recipient. I want you to construct a table with the following columns: the letter link, the original information I gave you, and the sender's location (sender_location, e.g., Canada, Ontario, South Woodslee). If there is no information just add ”unknown”. Do not include extra information like dates in the output.
  I will now give you the list of 50 letters, take as much time as you need, and provide the desired table. [Insert table]
- **Finding sender location using first lines of text when we don't know anything**
  I have a list of letters sent by Irish migrants (mostly in the US and Canada, but not only) to their family/friends in Ireland (though not always). This list includes the letter link, the receiver location (often missing) and then the first words of the letter.
  Those first words sometimes contain the location of the letter writer. However, we don't have any preprocessed information on the location of sender, so this might simply be missing from the first lines of text. I want you to construct a table with the following columns: the letter link, the original information on the receiver (receiver_info, return the exact same thing I gave you), and the sender's location (sender_location, e.g., Canada, Ontario, South Woodslee). If you couldn't find any information on the sender location, just add ”unknown”.
  Be careful not to confuse the sender location and the receiver location. The sender and the receiver cannot be in the same location.
  I will now give you the list of about 50 letters, take as much time as you need, and provide the desired table. [Insert table]
- **Refining broad location of receiver using first lines of text**
  I have a list of letters sent by Irish migrants (mostly in the US and Canada, but not only) to their family/friends in Ireland (though not always). This list includes the letter link at the beginning, some information on the location of the letter writer (sender_info), some broad information about the location of the letter recipient (receiver_info) and then the first words of the letter.

Usually, those first words contain the location of the letter writer and of the letter recipient. For each receiver's location, I want you to search in those first words if you can find a more precise location that is consistent with the broader one. For example, if the broad receiver location is "Pennsylvania", the more precise one has to be in Pennsylvania. If you don't find a location that respects this rule, just put "Unknown".

The output must be a table with the following 4 columns: the letter link, the original information I gave you on the location of the sender and on the location of the receiver, and a more precise receiver's location (receiver_location). Do not return the text in the output.

I will now give you the list of 50 letters. Do not say anything, take as much time as you need, and provide the desired table. [Insert table]

- **Finding receiver location using first lines of text when we don't know anything**

  I have a list of letters sent by Irish migrants (mostly in the US and Canada, but not only) to their family/friends in Ireland (though not always). This list includes the letter link, the sender location (sometimes missing) and then the first words of the letter.

  Those first words sometimes contain the location of the letter writer and of the receiver. However, we don't have any preprocessed information on the location of receiver, so this might simply be missing from the first lines of text. I want you to construct a table with the following columns: the letter link, the original information on the sender (sender_info, return the exact same thing I gave you), and the receiver's location (receiver_location, e.g., Canada, Ontario, South Woodslee). If you couldn't find any information on the receiver location, just add "unknown".

  I want you to be very careful not to confuse the sender location and the receiver location. I will give you two additional rules that you must respect:

  1. The sender and the receiver cannot be in the same location. For example, if you know from the information I gave you that the sender is in Ontario, it is impossible that the receiver is also in Ontario so you should take this into account in your answer. 2. Sometimes, in the first lines of text, there is a location preceding a precise date. When this is the case, the location is the sender's one, not the receiver's one, so you can ignore it.

  I will now give you the list of about 50 letters, take as much time as you need, and provide the desired table. [Insert table]

- **Finding receiver location using last lines of text when we don't know anything**

  I have a list of letters sent by migrants (mostly in the US and Canada, but not only) to their family/friends in Ireland (usually but not always). This list includes the letter link, the sender location (sometimes missing) and then the last words of the letter.

  Those last words sometimes contain the location of the letter writer and of the receiver. However, we don't have any information on the location of receiver, so this might simply

be missing from the last lines of text. I want you to construct a table with the following columns: the letter link, the original information on the sender (sender_info, return the EXACT information I gave you), and the receiver's location (receiver_location, e.g., Canada, Ontario, South Woodslee). If you couldn't find any information on the receiver location in the last lines of text, just add "unknown".
There is one rule that you must respect at any cost: the sender and the receiver cannot be in the same location. For example, if you know from the information I gave you that the sender is in California, you cannot return a location in California as the receiver's location.
I will now give you the list of about 50 letters. Take as much time as you need, and provide the desired table. Do not say anything else, just provide the table. [Insert table]

References

Abramitzky, R. and L. Boustan (2017). “Immigration in American economic history”. In: *Journal of Economic Literature* 55(4), pp. 1311–1345.

Abramitzky, R., L. Boustan, et al. (2020). “Do immigrants assimilate more slowly today than in the past?” In: *American Economic Review: Insights* 2(1), pp. 125–141.

Abramitzky, R., L. Boustan, et al. (2021a). “Automated linking of historical data”. In: *Journal of Economic Literature* 59(3), pp. 865–918.

Abramitzky, R., L. Boustan, et al. (2021b). “Intergenerational mobility of immigrants in the United States over two centuries”. In: *American Economic Review* 111(2), pp. 580–608.

Alroey, G. (2011). *Bread to Eat and Clothes to Wear: Letters from Jewish Migrants in the Early Twentieth Century*. Wayne State University Press: Detroit, Mich. 228 pp.

Alsan, M., K. Eriksson, et al. (2020). *Understanding the Success of the Know Nothing Party*. NBER Working Paper Series 28078.

Amador-Moreno, C. P. (2016). “Female Voices in the Context of Irish Emigration: A Linguistic Analysis of Gender Differences in Private Correspondence”. In: *Irish Journal of Applied Social Studies* 16(1).

Avila-Ledesma, N. E. (2019). ““Believe My Word Dear Father That You Can’t Pick Up Money Here as Quick as the People at Home Thinks It”: Exploring Migration Experiences in Irish Emigrants’ Letters”. In: *Corpus Pragmatics* 3(2), pp. 101–121.

Avila-Ledesma, N. E. and C. P. Amador-Moreno (2016). ““The More Please [Places] I See the More I Think of Home”: On Gendered Discourse of Irishness and Migration Experiences”. In: *Yearbook of Corpus Linguistics and Pragmatics 2016: Global Implications for Society and Education in the Networked Age*. Ed. by J. Romero-Trillo. Springer International Publishing: Cham, pp. 85–105.

Bandiera, O., M. Mohnen, et al. (2019). “Nation-building through compulsory schooling during the age of mass migration”. In: *The Economic Journal* 129(617), pp. 62–109.

Barsbai, T., H. Rapoport, et al. (2017). “The Effect of Labor Migration on the Diffusion of Democracy: Evidence from a Former Soviet Republic”. In: *American Economic Journal: Applied Economics* 9(3), pp. 36–69.

Barton, H. A. (2012). *Letters from the Promised Land: Swedes in America, 1840-1914*. 4th print. Published by the University of Minnesota Press for the Swedish Pioneer Historical Society: Minneapolis.

Bazzi, S., A. Ferrara, et al. (2023). “The other great migration: Southern whites and the new right”. In: *The Quarterly Journal of Economics* 138(3), pp. 1577–1647.

Beach, B. and W. W. Hanlon (2023). “Culture and the historical fertility transition”. In: *The Review of Economic Studies* 90(4), pp. 1669–1700.

Beine, M., F. Docquier, et al. (2013). “International Migration, Transfer of Norms and Home Country Fertility”. In: *Canadian Journal of Economics/Revue canadienne d'économique* 46(4), pp. 1406–1430.

Bertoli, S., V. Dequiedt, et al. (2016). “Can Selective Immigration Policies Reduce Migrants' Quality?” In: *Journal of Development Economics* 119, pp. 100–109.

Borges, M. J. (2016). “For the Good of the Family: Migratory Strategies and Affective Language in Portuguese Migrant Letters, 1870s–1920s”. In: *The History of the Family* 21(3), pp. 368–397.

Clarkson, L. A., L. Kennedy, et al. (1997). *Database of Irish Historical Statistics : Emigration, 1851-1911*. SN: 3580. UK Data Service.

Collins, W. J. and A. Zimran (2019). “The Economic Assimilation of Irish Famine Migrants to the United States”. In: *Explorations in Economic History* 74, p. 101302.

Connolly, S. (2022). *On Every Tide: The Making and Remaking of the Irish World*. Little, Brown Book Group.

Diabate, I. and S. Mesplé-Somps (2019). “Female genital mutilation and migration in Mali: do return migrants transfer social norms?” In: *Journal of Population Economics* 32(4), pp. 1125–1170.

Docquier, F., E. Lodigiani, et al. (2016). “Emigration and Democracy”. In: *Journal of Development Economics* 120, pp. 209–223.

Docquier, F., A. Tansel, et al. (2020). “Do emigrants self-select along cultural traits? Evidence from the MENA countries”. In: *International Migration Review* 54(2), pp. 388–422.

Dounia, M. (2020). “Transnational Practices and Emotional Belonging among Early 20th-Century Greek Migrants in the United States”. In: *Genealogy* 4(3), p. 90.

Droller, F. (2018). “Migration, population composition and long run economic development: Evidence from settlements in the pampas”. In: *The Economic Journal* 128(614), pp. 2321–2352.

Eyford, R. C. (2015). ““Close Together, Though Miles and Miles Apart”: Family, Distance, and Emotion in the Letters of the Taylor Sisters, 1881-1921”. In: *Histoire sociale/Social history* 47(96), pp. 67–86.

Fernández, M. (forthcoming). “Mass Emigration and Human Capital over a Century: Evidence from the Galician Diaspora”. In: *Journal of the European Economic Association*.

Ferrie, J. P. (1997). “The Entry into the U.S. Labor Market of Antebellum European Immigrants, 1840–1860”. In: *Explorations in Economic History* 34, pp. 295–330.

Fitzpatrick, D. (1980). “Irish Emigration in the Later Nineteenth Century”. In: *Irish Historical Studies* 22(86), pp. 126–143. JSTOR: 30008770.

Fitzpatrick, D. (2006). “Irish Emigration and the Art of Letter-Writing”. In: *Letters across Borders: The Epistolary Practices of International Migrants*. Springer, pp. 97–106.

Fouka, V. (2019). “How Do Immigrants Respond to Discrimination? The Case of Germans in the US During World War I”. In: *American Political Science Review* 113(2), pp. 405–422.

Fouka, V. (2020). “Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I”. In: *Review of Economic Studies* 87(1), pp. 204–239.

Fouka, V., S. Mazumder, et al. (2022). “From immigrants to Americans: Race and assimilation during the Great Migration”. In: *The Review of Economic Studies* 89(2), pp. 811–842.

Gabaccia, D. R. (2000). *Italy's Many Diasporas*. 1. publ. Global Diasporas. Univ. of Washington Press: Seattle. 264 pp.

Gerber, D. A. (2006). *Authors of Their Lives: The Personal Correspondence of British Immigrants to North America in the Nineteenth Century*. New York Univ. Press: New York. 422 pp.

Giuliano, P. and M. Tabellini (2020). “The seeds of ideology: Historical immigration and political preferences in the United States”. In: *NBER Working Paper Series 27238*.

Gráda, C. Ó. (2019). “The Next World and the New World: Relief, Migration, and the Great Irish Famine”. In: *The Journal of Economic History* 79(2), pp. 319–355.

Gráda, C. Ó. and K. H. O'Rourke (1997). “Migration as Disaster Relief: Lessons from the Great Irish Famine”. In: *European Review of Economic History* 1(1), pp. 3–25.

Hatton, T. (2001). “The Age of Mass Migration: What We Can and Can't Explain”. In: *Migration and Mobility: The European Context*. Ed. by S. Ghatak and A. Showstack Sassoon. Palgrave Macmillan UK: London, pp. 11–29.

Hirota, H. (2016). *Expelling the Poor: Atlantic Seaboard States and the Nineteenth-Century Origins of American Immigration Policy*. Oxford University Press.

Hu, M. and B. Liu (2004). “Mining and Summarizing Customer Reviews”. In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Kdd '04. Association for Computing Machinery: New York, NY, USA, pp. 168–177.

Hurlburt, S. (2017). “Kin Keeping and Family Storytelling in Nineteenth-century French-Canadian Immigrant Letters: The Bergevin Corpus”. In: *Histoire sociale/Social history* 50(102), pp. 285–313.

Ishiguro, L. M. (2011). “Relative Distances: Family and Empire between Britain, British Columbia and India, 1858-1901”. PhD thesis. University College London.

Jaro, M. A. (1989). “Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida”. In: *Journal of the American Statistical Association* 84(406), pp. 414–420.

Jaschke, P., S. Sardoschau, et al. (2022). *Scared Straight? Threat and Assimilation of Refugees in Germany*. NBER Working Paper Series 39381.

Kampanelis, S., A. Elizalde, et al. (2023). “Songlines”. In: *SSRN working paper*.

Karadja, M. and E. Prawitz (2019). “Exit, voice, and political change: Evidence from Swedish mass migration to the United States”. In: *Journal of Political Economy* 127(4), pp. 1864–1925.

Kerby, E., A. Moradi, et al. (2022). “African time travellers: What can we learn from 500 years of written accounts?” In: *The Economic History Review*.

Knudsen, A. S. B. (2024). “Those who stayed: Individualism, self-selection and cultural change during the age of mass migration”. In: *mimeo*.

Levitt, P. (1998). “Social Remittances: Migration Driven Local-Level Forms of Cultural Diffusion”. In: *International Migration Review* 32(4), pp. 926–948.

Levitt, P. and D. Lamba-Nieves (2011). “Social Remittances Revisited”. In: *Journal of Ethnic and Migration Studies* 37(1), pp. 1–22.

Lodigiani, E. and S. Salomone (2020). “Migration-induced transfers of norms: the case of female political empowerment”. In: *Journal of Demographic Economics* 86(4), pp. 435–477.

Manacorda, M., J. Ponticelli, et al. (2024). “The International Transmission of Democratic Values: Evidence from the African Diaspora”. In: *mimeo*.

Markelis, D. (2006). ““Every Person like a Letter”: The Importance of Correspondence in Lithuanian Immigrant Life”. In: *Letters across Borders: The Epistolary Practices of International Migrants*. Springer, pp. 107–123.

Melki, M., H. Rapoport, et al. (2024). *Cultural Remittances and Modern Fertility*. NBER Working Paper Series 32990.

Michalopoulos, S., E. Papaioannou, et al. (2024). “Songs of Refugees”. In: *mimeo*.

Michalopoulos, S. and C. Rauh (2024). *Movies*. Tech. rep. NBER Working Paper Series 32220.

Michalopoulos, S. and M. M. Xue (2021). “Folklore”. In: *The Quarterly Journal of Economics* 136(4), pp. 1993–2046.

Miho, A., A. Jarotschkin, et al. (2024). “Diffusion of gender norms: Evidence from Stalin’s ethnic deportations”. In: *Journal of the European Economic Association* 22(2), pp. 475–527.

Miller, K. A. (1985). *Emigrants and Exiles: Ireland and the Irish Exodus to North America*. 1st ed. Oxford University Press Inc.

Miller, K. A. and B. D. Boling (1991). “Golden Streets, Bitter Tears: The Irish Image of America during the Era of Mass Migration”. In: *Journal of American Ethnic History* 10 (1/2,), pp. 16–35. JSTOR: 27500798.

Miller, K. A., A. Schrier, et al. (2003). *Irish Immigrants in the Land of Canaan: Letters and Memoirs From Colonial and Revolutionary America, 1675–1815*. Oxford University Press.

Moreton, E. (2015). ““I Hope You Will Write”: The Function of Projection Structures in a Corpus of Nineteenth-Century Irish Emigrant Correspondence”. In: *Journal of Historical Pragmatics* 16(2), pp. 277–303.

Munch, L. N. (2021). *Hilsen fra Amerika: historien om de danske udvandrere og deres postkort, billeder og breve*. Turbine.

Ooms, J. (2024). *Tesseract: Open Source OCR Engine*. Version 5.2.1.

Paul-Dubois, L. (1908). *Contemporary Ireland*.

Posch, M. (2024). “Cultural Change: Evidence from Three Centuries of U.S. Local Newspapers”. In: *mimeo*.

Rocha, R., C. Ferraz, et al. (2017). "Human capital persistence and development". In: *American Economic Journal: Applied Economics* 9(4), pp. 105–136.

Rosenthal, C. J. (1985). "Kinkeeping in the Familial Division of Labor". In: *Journal of Marriage and the Family* 47(4), p. 965. JSTOR: 352340.

Ruggles, S., M. A. Nelson, et al. (2024). *IPUMS Ancestry Full Count Data: Version 4.0*. Minneapolis, MN: IPUMS.

Sequeira, S., N. Nunn, et al. (2020). "Immigrants and the Making of America". In: *The Review of Economic Studies* 87(1), pp. 382–419.

Spilimbergo, A. (2009). "Democracy and Foreign Education". In: *American Economic Review* 99(1), pp. 528–543.

Tabellini, M. (2020). "Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration". In: *The Review of Economic Studies* 87(1), pp. 454–486.

Tuccio, M. and J. Wahba (2018). "Return Migration and the Transfer of Gender Norms: Evidence from the Middle East". In: *Journal of Comparative Economics* 46(4), pp. 1006–1029.

Tuccio, M. and J. Wahba (2020). *Social remittances*. Springer.

Voth, H.-J. and D. Yanagizawa-Drott (2024). "Images". In: *mimeo*.

Zempel, S. (2000). *In Their Own Words: Letters from Norwegian Immigrants*.

Concluding Remarks

This concluding chapter summarizes the main takeaways of this dissertation. It addresses some limitations of this work and briefly discusses some of the avenues for future research.

In the first chapter, we explored whether attitudes towards immigration affect the migration plans of individuals who still live in their home countries. We found that more hostility towards migrants in a country leads to a decrease in the number of people planning to migrate to this country. Interestingly, we observe similar effects across all skill levels.

In the second chapter, I investigated a similar question but focused on the effect of having a populist leader at power on international migration flows. I found that right-wing populist leaders in particular lead to a decrease in immigration growth rates, with a larger negative effect on low-skilled migration flows. However, I didn't find any effect on emigration growth rates, suggesting that the "voting with the feet" mechanism doesn't play a role here. When investigating the mechanism behind those results, I found that the main driver is the implementation of stricter migration policies that make it harder to come in the country or that lower the economic integration of migrants.

Finally, the third chapter addressed a different question, namely how do migrants diffuse culture, including political norms and values, across countries? We performed some analysis on the content of the letters and we showed that gender plays an important role, as the importance of economics and politics is much lower in letters written by women. We also found that having a larger Irish diaspora close to the letter writer has a negative effect on the amount of social remittances sent, suggesting that migrants tend to have less ties with their homeland as their local network becomes more important.

Taken together, those chapters give several insights on the relationship between international migration and the creation and diffusion of all sorts of political preferences and attitudes. The first two chapters have shown that hostility towards immigration, either in terms of attitudes only or in terms of implementing more restrictive immigration policies, do affect migration flows. Importantly, this affects migrants of all skill level and origins. Therefore, it goes beyond the now frequent discourse about reducing migration from poorer, more culturally distant countries. This can also have detrimental economic effects on the destination country, as migration could for instance soften the fiscal burden of aging (Bernardino et al., 2024)[52]. The insights drawn from the third chapter are different. The main one is that one should not assume that all migrants send the same kind of social remittances (including information about the norms and values of the country they live in, as well as information on the level of discrimination they face). The content of migrants' communications varies importantly depending on the individual characteristics of migrants, but also due to the socio-economic environment they live in.

The results coming from this dissertation should nonetheless be nuanced. Each chapter focused on different contexts: the first one looked at migration towards Europe in the last 15 years, the second one looked at populist leaders and migration flows in the world in the last 60 years, and the third one focused on Irish migrants in North America in the period 1840-1930. As shown by the vast literature on migration economics, it is hard to draw conclusions that apply to all contexts and as such, those results should be taken with caution. More research is also needed in this area, notably to explore further whether anti-immigration attitudes could lead to a vicious circle by causing negative selection of migrants in terms of skills. While we didn't find this mechanism in the first chapter, studies at a finer geographical level could provide different findings. Regarding the last chapter on social remittances, recent efforts to digitize more collections of letters in various countries have paved the way for more research on this topic, which is an essential component of international migration.

References

Bernardino, T., F. Franco, et al. (2024). *The Costs of Building Walls: Immigration and the Fiscal Burden of Aging in Europe*.

Eurostat (2023). *Ageing Europe - Statistics on Population Developments*.

[52] According to Eurostat, the ratio of the 65+ over 20-64-year-old population will increase by more than 20 percentage points, from 34.1% in 2019 to 56.7 percent in 2050, significantly affecting public finances (Eurostat, 2023).