Université catholique de Louvain
Economics School of Louvain

Political history, migration and economic outcomes

Charles de Pierpont de Burnot

June 16, 2026

Composition du Jury:

Mario Larch	(University of Bayruth)
Fabio Mariani	(Université catholique de Louvain)
Luca Pensieroso	(Université catholique de Louvain)
Hillel Rapoport	(Paris School of Economics)
Gonzague Vannoorenberghe (Promoteur)	(Université catholique de Louvain)

Président du Jury:

Luca Pensieroso (Université catholique de Louvain)

Acknowledgements

Many people deserve my most grateful sentiments for their support throughout this PhD journey, yet my words seem insufficient to honour them in a manner that truly does justice to their benevolence and selfless dedication to others. I therefore wish to borrow a few words from minds far greater than my own, whose works accompanied me throughout this doctoral journey, in the hope that their eloquence may lend support to the gratitude I struggle to express appropriately.

As my strongest and most reliable support throughout this journey, both professionally and morally, my first, and most extensive, though still inadequate, expression of gratitude is for Gonzague Vannoorenberghe, who largely exceeds what is commonly expected from an advizor in terms of duties. I am grateful and astonished by all the efforts he made for me. To express my admiration and make the reader aware of his patience toward me and my numerous eccentric ideas, I'll use a small story from the Viscount of Bragelonne (A. Dumas) *where I'll grant Gonzague a crown: D'Artagnan, fulfilling his idea to kidnap and offer the general Monck to the king Charles of England after numerous efforts and adventures, finally brings his prey to the king at an unexpected moment. Then the king patiently explained to him that this was unnecessary and did little to advance the greater plan, while remaining supportive and amused by his earnest efforts. This scene was repeated so many times at the office of Gonzague, that I believe that we could outperform professional actors in a cinematic adaptation of the book for this specific act. Gonzague did far more than merely play a role, though: He helped me grow as an economist with great patience, kindness, and benevolence.*

As members of my jury, I want to thank Luca Pensieroso, Fabio Mariani, Mario Larch, and Hillel Rapoport for their guidance, comments, and kind words during the private defense. I learned many things from each of them, in different ways that I do not resist describing.

"What is this animal?" *- said Luca Pensieroso in 2020, addressing an over-sophisticated equation on the blackboard in front of his students. Through lessons, Luca and Fabio taught me the genealogy of our science and the fundamentals of macroeconomics at a time when I could scarcely even pronounce "heteroscedasticity" correctly. While I was initially attracted by purely applied economics with a problem-solving mindset, they managed to increase my interest in the art of theoretical modeling and my curiosity toward all the related animals. From teachers, they became colleagues as I started my PhD, and through regular interaction and shared passion for both economics and history, friends who were always ready to provide me with valuable feedback that greatly helped in the development of the present contribution.*

"The Ancients left us models of epic poems in which all interest is concentrated on the heroes. Since these poems are masterpieces, we persist in failing to see how outdated such a conception of history has become in our own age." *- wrote Count Tolstoy in* War and Peace *while precisely making a review and criticism of a specific period in Russia using a poetic story close to the ones of the "Ancients". Hillel impressed me in much the same way, through his ability to transform economics into a poetic science with both elegant and founded narratives. Our regular reunions, his academic rigor, and his valuable comments greatly improved the readability and academic quality of my works. I also thank Hillel for hosting me during my stay at the* Paris School of Economics*, where he made great efforts to integrate me into the research community and took the time to discuss my latest research with me on several occasions.*

"Gravity is endless fun!" *- This quote is attributed to Peter Neary by Yoto V. Yotov in (one of) his trade gravity toolkit papers, a statement that they both share as he concludes by "He was right!". While I met Mario during my last year of PhD, I already "knew" him before through his countless contributions to the gravity literature, bringing more fun to this world. His expertise and highly complete recommendations were a great help in the last few months, but I also want to thank him for one of his specific works that greatly inspired the first chapter of this thesis.*

"Generosity, humanity, kindness, compassion, mutual friendship and esteem, all the social and benevolent affections, when expressed in the countenance or behaviour, [...] please the indifferent spectator upon almost every occasion." *- By repeating those lines from the Adam of the Economics Genesis, I wanted to thank all the economists and colleagues I interacted with along the way at UCLouvain, and more particularly, the professors, as well as my solid PhD peers who lost a fraction of their life* *drinking coffee with me: Adam Levai, Akira Momota, Alain Guillet, Anna Queck, Amandine Magis, Angel Pandit, Antoine Germain, Aurélie Bertrand, Arastou Khatibi, Arnaud Deseau, Arnaud Beaurin, Bastien Latour, Brieuc Pierre, Bruno Van der Linden, Catherine Dewit, Catherine Powis de Tenbossche, Carlo Ciccarelli, Chiara Zanardello, Chun Che Kok, Cyril Ghislain, Daniele Verdini, David de la Croix, Diego Malo Rico, Dorothee Hillrichs, Ella Sargsyan, Elsa Leromain, Esmeralda Gerritse, Engie Theys, Fabio Blasutto, Félix Dopchie, Filippo Manfredini, Flora Sarigul, Florian De Bundel, Francesco Pascucci, Francesco and Sarah Roccazzella, François Courtois, Gabriel El Mahi, Gaia Spolverini, Géraldine Carette, Gianmarco Luu, Giulia Tarullo, Hakimeh Hemmatipour, Jacques Cartuyvels, Jean Hindriks, Jean-François Maydstadt, Jing-Rong Zeng, Karine Boukkadem, Laura Angelini, Leïla Van Keirlbilck, Luigi Boggian, Maïlys de Bernard de Fauconval, Marcus Biermann, Margaux Clarr, Marie Gilot, Marie Gonze, Marouane Masnaoui, Mathilde Pourtois, Mathilde Sage, Martial Toniotti, Matthis Dartevelle, Mélissa Motin, Muriel Dejemeppe, Natalia Bermundez, Nathan Lachapelle, Nicolas Stevens, Niranjana Prasad, Quentin Lété, Ricardo Turati, Riccardo Paba, Roberto Sanchez Torres, Robin Philips, Robin Schindowski, Rossana Scebba, Sébastien Schillings, Sébastien Van Bellegem, Serena Chellini, Séverine Dinjar, Silvia Peracchi, Tanguy Isaac, Thomas Eisfeld, Thomas Lebbe, Tiziano Toniolo, Valérie Pottie, Vincent Bodart, Vincent Notte, Vincent Vandenberghe, WilliamParienté, and Zhaniya Idrissova. These persons possess so many qualities that they would let no spectators remain indifferent.*

"I shall forge Gaul into a single vector of wills, [...] before which nothing shall stand." *- I had the pleasure to read these words, attributed by* Iulius Caesar *to* Vercingetorix *in* The War of Gaul, *during my visiting at the* Paris School of Economics *in early 2025, where common agreement on the right method to apply for identification would require at least the intervention of the famous Celtic king. The dynamism of our discussions, the passion towards the truth and academic excellence, our late shared meals and beers, the numerous seminars, and the constant chaotic atmosphere of the official PhD WhatsApp group made this visit a vector of both great friendships and academic growth. I want to thank all the amazing people I learned from during my stay: Anne-Célia Disdier, Antonio Léon Fernandez, Arthur Obminski, Ariell Reshef, Dilek Sevinc, Fabio Traverso, Filipp Bothe, Gemma Harris, Hector Paredes Castro, Julia Paul-Venturine, Katrin Millock, Leo and Nolwenn Czaika, Maëlle Stricot, Matthieu Parenti, Marcelo Gantier, Manpreet Singh, Miko Lepistö, Louis-Marie de Belleville, Patrick Petit, Pierre Brassac, Rafael Schütz, Sandra Poncet, Thierry Verdier, Valentina Gabrielli, Vivien Liu, and Youssef Salib. A particular thanks to the GPET tribe/research group that welcomed me.*

"Ah!" *- said Donatien Baudelet, a friend who cheered me along the way and who would be the most perfect substitute of the presentator Denis Broignart, when he passed randomly at my office. I would be remiss not to thank those who made this thesis possible, as well as the friends who supported me throughout this journey: Hakadoch Baroukh Hou, Marie De Wit for her unconditional support and affection, Gatien Cort van der Linden for sharing a glass of whiskey with me whenever I needed it, Timo Tallgren, who shares my linguistic and reading passions, Quentin Depreter, my anchor through stormy seas, Harold della Faille and Eloïse van Zeeland for their patience in enduring the eccentricities that inevitably accompany academic life, Axel Laniez, who sent me countless of messages to grab news about the thesis, Charles-Hubert and Marguerite-Marie Clérin and their children Léon-Hubert and Charles-Foucaut, my second family, Malo and Apolline Béguin and their chidren Léopold and Camille, and Thibault de Pierpont for his constant and contagious optimism.*

In a thesis studying deeply long-run dynamics, I wanted to conclude the acknowledgment by thanking the long-run roots of most of my achievements: My parents, Albert and Nathalie de Pierpont. I thank them not only for their early investment in my education, but mostly for their love and constant support during my Odyssey and beyond. A special thought also goes to my grandmother, Bernadette de Pierpont, whose kindness is matched only by her smile. I also wish to remember my grandfather, Dieudonné, who left us nearly twenty years ago and whose memory remains cherished.

Contents

List of Figures

List of Tables

General Introduction

This thesis studies the economic consequences of globalization. Arguing about globalization and its influence on economic outcomes or linkages sounds considerably cloudy, at least without providing a clear, restricted definition of this concept. While several academics have attempted to define globalization, the diversity of the existing definitions is proof that a definition of globalization is highly contextual and evolves according to the ambition of a research agenda and its scale. This thesis is no exception to this rule, and I express here a definition that only embodies a fragment of what globalization may reflect: increasing political integration, cultural exchanges, as well as goods and people's geographic mobility. The subsequent increase in connections between human communities is likely to have a strong influence on local economic outcomes and ties. This thesis attempts to evaluate the importance of these phenomena by exploiting historical migration and political integration episodes in Europe and the United States of America.

The chapter 1 investigates the consequences of past political integration on the long-run economic linkages between human communities. It uses Western Europe as a case study, exploiting its well-documented early integration during the Roman era. Political integration is usually known to affect trade through industrial specialization related either to past planned economies or decreasing trade costs. While cultural traits are well known to affect economic ties, they are often regarded as exogenous, almost *God-given* characteristics. But cultures change, especially in a globalized world, and their differences have roots. This chapter argues that an important part of the Western European cultural fragmentation and economic ties between European regions arises from the European political history. More specifically, it provides evidence of the long-run influence of one of the very early main globalization shocks that Western Europe experienced: the Roman conquest and its subsequent institutions.

The chapter 2 turns to a more contemporary manifestation of globalization by examining the local labor-market consequences of immigrant birthplace diversity in the United

States. A growing empirical literature documents weakly positive or null average effects of regional diversity on wages. This chapter argues that such findings may reflect substantial heterogeneity driven by aggregation. Adopting a granular approach, it shows that diversity among workers within the same metropolitan area, sector, and occupation is positively associated with wages, particularly when individuals share similar levels of education and experience. By relying on a shift-share instrumental strategy and exploiting the detailed composition of migration inflows during the early twenty-first century, the chapter disentangles complementarities arising from skill similarity and interaction intensity from broader national specialization patterns. It further investigates the influence of economic and cultural distances on productivity gains related to diversity: complementarities are strongest when economically and culturally proximate groups interact within occupations, while culturally distant origins enhance gains among workers with similar educational backgrounds. These findings suggest that productivity effects operate through finely structured local interactions rather than broad regional diversity measures.

The chapter 3 revisits the debate on the labor-market consequences of migration shocks for native workers. Although the long-run neutrality of migration on average native wages and employment is widely accepted, short- and medium-run empirical estimates remain strikingly heterogeneous. This chapter argues that such divergence largely reflects differences in how migration flows, but also spatial and occupational mobility, are incorporated into empirical designs. More specifically, it revisits estimates using a shift-share approach, one of the most common identification tools in this literature, using a single U.S. dataset while systematically varying the level of analysis. Accounting for a bias arising from serially correlated migration flows, it reconciles different strands of the literature by showing a convergence in the sign of the results: a negative short-run impact of migrations on wages in the US. Additionally, this chapter documents the poor performance of the shift-share instrument at aggregated levels, contributing to the literature on its evaluation.

Chapter 1

The Roman roots of the Western European trade

This chapter examines the persistent legacy of Roman integration in current trade linkages across Western Europe. Applying restrictive specifications to a sample of 16,256 region pairs, as well as a spatial regression discontinuity design to the North Roman Limes, it provides evidence that regions jointly integrated within the Roman states (extensively or for a longer period) trade significantly more with each other today. Part of this influence reflects subsequent political history, the current transport networks, which sometimes follow past Roman roads, as well as geographic features that may have induced trade without the Roman intervention, such as resource differences and geographic connectivity. However, a remaining sizable part of the estimated net effect of the Roman rule is attributable to cultural convergence measured by preference and linguistic similarities. These findings highlight a long-lasting influence of past political integration on the European market integration.

1.1 Introduction

Does political integration shape long-run economic linkages? Despite the removal of major barriers to trade, parts of bilateral trade patterns appear relatively persistent beyond geography and contemporary institutions (Botsas, 1987; Ward and Hoff, 2007; Head and Mayer, 2013; Santamaria et al., 2023), notably in Western Europe, where demographic similarities and network play an important role (Head and Mayer, 2000; Disdier and Mayer, 2007; Chaney, 2014). These factors are likely to have historical roots (Becker et al., 2020). This chapter argues that part of contemporary European economic integration reflects a much older process: past episodes of deep political integration that have generated durable interregional economic linkages through historical connections and cultural convergence.

More particularly, this research assesses the remaining influence of the Roman political integration on current trade linkages. In the Western European context, the Roman states indeed represent the very first large-scale, and historically most extensive political integration in terms of both size and duration[1]. They provided institutions and infrastructure that deeply affected ties between regions in the imperial dominion: Historical evidence indicates a strong increase in trade during the Roman era[2] as well as the diffusion of new ideas and consumption habits[3]. To the best of my knowledge, it is the first research documenting the long-run influence of the Romans through political integration and institutions on current actual trade at a broad Western scale[4].

[1] The fig. 1.1 depicts the historical evolution of the dominions of the widest Western European states in Western Europe, highlighting how the Roman states compare to later political integration.

[2] See among others Hopkins (1980), pp. 105-06, or Wilson and Bowman (2017), pp. 16-20. The section 1.B.5 fig. 1.B8 map depicts the Roman-era Western *Terra Sigillata* trade flows, which support a relatively strong trade integration during that period.

[3] See Beard (2022), Chapter XII, for cultural changes under the Roman rule and the section 1.B.4 for evidence of the influence of the Roman rule and transport infrastructure on the diffusion of Christianity.

[4] It is worth noting that such persistent influence, if it exists, may have been attenuated or amplified by important barriers to Mediterranean trade arising in the 8th-9th centuries: Following the Arab conquest of North Africa and the Middle East, there was indeed a displacement of the European core from the Mediterranean to Western and central Europe, as well as a progressive cultural shift. While numerous findings of Arab coins in Western Europe may suggest that it did not strongly impact trade between the Arab and the Western worlds (Persson and Sharp, 2015, pp.14-18), recent evidence shows a strong trade disruption (Boehm and Chaney, 2024).

It belongs to a literature assessing the long-lasting impact of the Romans' decisions on the regional economic development and economic ties. While the Roman contribution to contemporary local economic performance became recently relatively well documented (Wahl, 2017; Dalgaard et al., 2022; Fritsch et al., 2023), little has been done on the persistence of economic linkages following this integration. A notable exception and closest work to the present research is from Flückiger et al. (2022): They detect a continuous historical influence of the early Roman transport network infrastructure on economic ties, until recently on the 2018 financial linkages. Their results also show that the Roman connectivity is associated with later stronger mobility (outside road networks), industrial specialization, and cultural proximity. The present chapter differs by looking at the Roman political integration beyond connectivity and infrastructure. It instead considers current trade linkages as the main outcome variable[5]. By doing so, it belongs to a literature assessing the determinants of trade[6] by exploring empirically its historical roots[7]. More broadly, this chapter also contributes to a literature exploring the long-lasting economic consequences of historical political fragmentation (Jones, 2003; Mokyr, 2016; Dincecco and Onorato, 2016; Blaydes and Paik, 2021; Delabastita et al., 2026).

To explore the relationship between current trade intensity and past Roman political integration, this article uses a sample of 16,256 Western European regional pairs[8] measuring

[5]Financial and trade linkages may differ, as investment and transport costs do not necessarily align; trade is indeed more sensitive to transport costs, which have been subject to strong asymmetric shocks throughout history (There were many shifts since the 15th century (Federico et al., 2021), especially in the 19th century, where Europe experienced major technological innovations that resulted in deep changes in trade flows, as enhanced by Pascali (2017)). They should have mitigated the Roman influence on trade if it merely reflected a past decrease in direct transport costs. A strong relationship between trade and financial ties is also not granted. Vertical financial linkages may entail trade, but horizontal financial linkages do not necessarily do so: For instance, a firm buying another that produces its inputs in another region is likely to import them at lower prices, generating trade, but a firm buying another that produces the same goods and services may actually make investment and trade substitutes (Lankhuizen et al., 2011). However, if cultural convergences and industrial divergences result from long-run integration, the Roman rule may have fostered trade linkages surviving the transport costs shocks beyond such considerations.

[6]See Santamaría et al. (2021), Anderson and Van Wincoop (2003), Feenstra (2002) or Havranek and Irsova (2017) among others.

[7]See Mazhikeyev and Edwards (2021) and Marczinek et al. (2025).

[8]the section 1.B.8 provides the detailed list in table 1.B19 and its geographic scale depicted in fig. 1.B12. To my knowledge, the NUTS2 regions are the most disaggregated level at which bilateral trade data could be retrieved at the time of the present study.

inter-regional trade intensity from European freight data in the period 2011-2019. It relies on a gravity framework and Poisson pseudo-maximum likelihood (PPML) estimators[9]. The inclusion of origin and destination fixed effects allows for the control of unobservables specific to the regions. It also uses various bilateral measures of geographic differences and similarities to disentangle the Roman influence from natural resources and geographic connectivity, as well as current trade costs. As there may be joint unobservables related to the geographic determinants of state formation or posterior history, I use two main strategies to approach causal identification: I rely on new historical geospatial data describing worldwide political history to disentangle the Roman influence from posterior events, decisions, and state formations, as well as on a spatial regression discontinuity design (RDD) strategy based on the Roman Limes that allows for comparing similar regions. In doing so, this paper attempts to shed light on the influence of past political integration on economic linkages, considering the effect of past institutions and isolating the Roman influence from posterior political history and geography.

The results suggest that the Roman joint political integration of Western European regions has continuously affected trade linkages until now; this stable influence seems notably associated with the heterogeneity in the contemporary *Border effect*, showing that it is driven by regions belonging to different current states that were strongly integrated during the Roman era. In that sense, Roman political integration appears as an attenuating factor in the impact of Western Europe's current fragmentation into different states on economic linkages. However, the Roman legacy seems to affect intranational trade patterns as well: For instance, a Western German region on the Roman side of the Limes trades 30% to 55% more with a neighbouring region on the same side than with a neighbouring region on the German side. The Roman influence decreases when one accounts for subsequent political integrations and road infrastructure, but a sizable part remains associated with cultural

[9]See Hillrichs and Vannoorenberghe (2022) on the trade gravity modeling in the presence of a *Home bias*. See Silva and Tenreyro (2006) as well as Head and Mayer (2014) on the interests of these estimators in the context of gravity frameworks that allow for including the extensive margins of trade in the analysis and account for heteroscedasticity.

proximity as measured by Latin linguistic roots and similarities in preferences. These results are consistent with a cultural imprint of long-past political integration (Grosjean, 2011; Grosfeld and Zhuravskaya, 2015) that may partially explain the highly documented and historically persistent cultural divide between and within countries (Giavazzi et al., 2019; Desmet and Wacziarg, 2021; Jackson and Medvedev, 2024), resulting in long-lasting economic linkages through the ages.

This article is structured as follows: The second section presents a literature review suggesting a continuous influence of political integration and institutions on economic linkages and the potential mechanisms behind the persistence of the Roman influence. The third section presents the data, while the fourth section presents the empirical methodology. The main results are displayed in a fifth section and potential mechanisms in a sixth.

1.2 Why Rome?

Past integration under Roman reign is a precursor to a greater economic performance. The literature emphasizes that regions in the imperial dominion are associated with higher current levels of innovation, entrepreneurship, well-being, and economic activities (Wahl, 2017; Dalgaard et al., 2022; Fritsch et al., 2023; Obschonka et al., 2025). Persistence in access to the road network the Romans built is identified as a key factor behind these findings. Roman rule would have fostered economic development through the foundation of towns and by connecting regions through the first (large-scale) Western European road network that existed. This long-lasting influence is particularly relevant for Western Europe[10] because many institutions in this area were Roman-rooted and relatively spatially uniform whereas the eastern part of the Empire was institutionally more heterogeneous, influenced by Greek

[10]In the context of this research, Western Europe is defined as Austria, Belgium, England, France, Germany, Italy, Luxembourg, Netherlands, Portugal, Spain, Liechtenstein, and Switzerland. The regions included are listed in the section 1.B.8 table 1.B19.

culture[11], and already had large-scale road networks before the Roman period. While the Roman contribution to European regional development is stated and well documented, much less is known about the underlying patterns of trade integration that may have sustained or transmitted these effects. This section first presents a brief literature review on the influence of political integration on trade, and second draws on insights from the economic and historical literature to discuss the potential mechanisms through which Roman political integration may have generated persistent patterns of trade integration.

1.2.1 A border effect from the past

It is well known that political integration has a positive impact on trade. In his seminal study on US-Canada inter-regional trade, McCallum (1995) shows that conditional on economic sizes and distance, goods are exchanged 22 times more within national borders than between borders. Since then, methodological advances have reduced this estimate to a single digit, in a range between 3 and 5 (Anderson and Van Wincoop, 2003; Feenstra, 2002; Havranek and Irsova, 2017). In the European context, Evans (2003) noticed a strong border effect between European countries before the trade and political integration of the EU members. Following this integration, intranational trade was still about ten times higher than international trade with an EU partner country. In the early 21st century, international trade between European regions remained about 5 times lower than their intranational trade (Santamaría et al., 2021; Frensch et al., 2023). This is likely to reflect much more than pure transportation costs, but also political, cultural, or linguistic differences, as well as home-biased consumer preferences. The heterogeneity among industries also suggests that industrial specialization is a core mechanism.

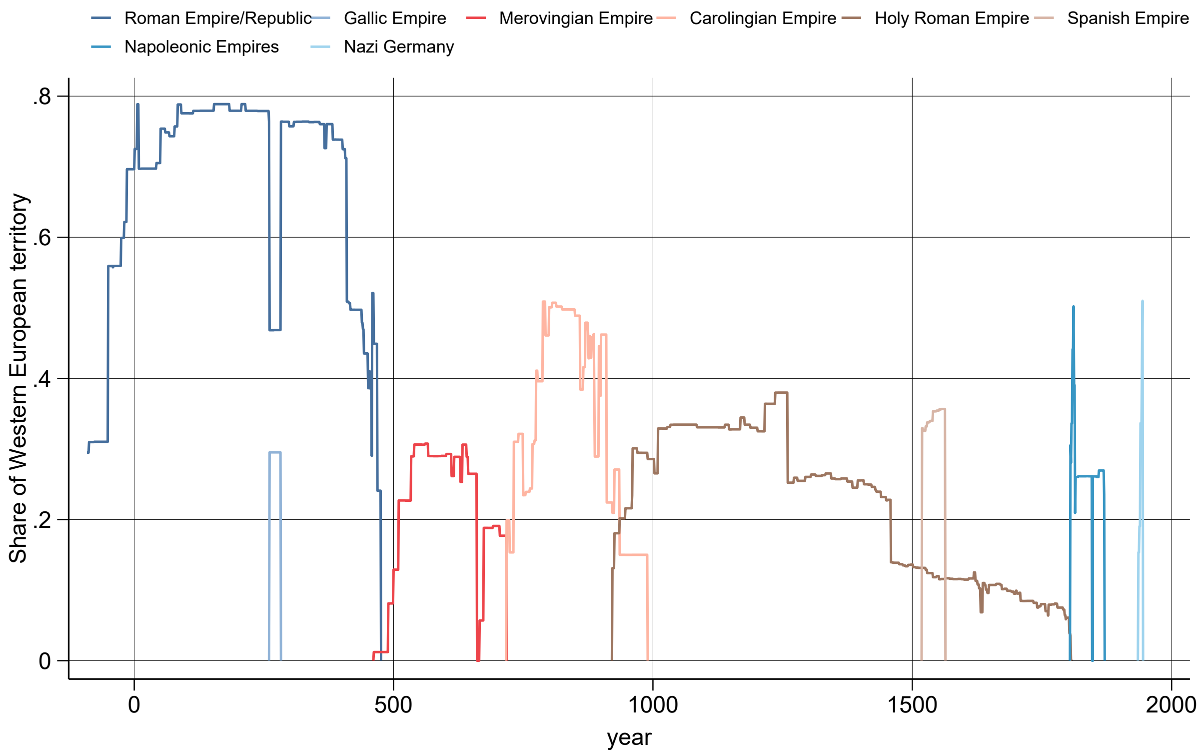
Recent articles emphasize that even past political integration influences economic linkages. Head et al. (2010) show that historical colonial linkages have a positive impact on trade in

[11] See Creanza (2024), Yon (2004), or Bekker-Nielsen (2006) pp.11-14, for examples of Greek influences shaping economic linkages and features. See Woolf (1994) on the relatively weaker Roman cultural influence in Eastern Europe than in Western Europe, or Lomas (2005) on how the cultural influence was mostly from the Greeks to the Romans rather than the opposite.

the late 20th century. The influence slowly decreased through time following independence, but never reached 0. Mazhikeyev and Edwards (2021) found that countries that were part of the USSR continued to trade more with each other decades after its collapse. These findings are robust to common languages and later trade agreements. The USSR was one of the biggest states that existed for more than half a century. Its planned economy is likely to have affected economic linkages until now (Djankov and Freund, 2002).

1.2.2 How Roman integration could affect trade in the long run

The Roman states differ from the previously mentioned cases: they were closer to a free economy (Temin, 2012) and existed in a far further era. Its persistent, continuous impact on economic linkages may therefore sound surprising: lower Roman-era transport costs between regions are associated with a faster onset of the Black Death plague in the 14th century, a higher correlation between 13th-17th-century crop prices, and stronger financial linkages in 2018 (Flückiger et al., 2022). These proxies are also strongly correlated with the duration of the common Roman rule of the Western regions. Why would the Romans matter? The long-run impact of the Roman settlements on development, making regions wealthier in the long run and more likely to trade, is not a sufficient answer: It is consistent with Ortman et al. (2024)'s findings on Roman coins in the United Kingdom suggesting that early urban settlement in England became trade center, but it is mechanically excluded from the gravity estimates of Flückiger et al. (2022). However, The Roman Empire represents the greatest Western European political integration in terms of size and time combined, as illustrated in fig. 1.1. Such a deep integration may have a longer influence than shorter, more recent ones. It is also associated with specific institutions that are known to have influenced the institutions of the later periods in Europe. The following paragraphs detail potential mechanisms behind this integration that may explain a long-lasting influence on economic linkages.



Note: The graph depicts the share of the presented states across time in the total Western Europe territory defined as the current Belgium, Germany, France, Spain, Italy, Portugal, United Kingdom, Liechtenstein, Andorra, and Switzerland. The graph only includes states that covered at least 25% of the territory. The Roman state is composed of the late Roman Republic, the Roman Empire, and the Western Roman Empire. The Gallic Empire was an internal Roman rebellion in 260 BC. The Holy Roman Empire included its union with the kingdom of Sicily in 1216. It is worth noting that the Spanish Empire and the Holy Roman Empire shared the same monarch while being technically different states. The Napoleonic Empires are composed of both the first and the second Empires. Source: Author's computation on *Cliopatra* (Bennett et al., 2025) and ESRI shapefiles.

Figure 1.1: Evolution of the shares of the biggest European states in the total Western European surface, 91 BC to 2024 AC

The persistence of Roman infrastructure

A first potential answer follows the findings of Flückiger et al. (2022): They mention that the Roman road network was the very first Western large-scale road network, and that it defined an important part of the current network. De Luca (2016) documents that the Roman roads were indeed still maintained and used during the European Middle Ages, indicating they should have affected trade linkages for a long period after the end of the Western Empire. Dalgaard et al. (2022) and De Benedictis et al. (2023) show that even a sizable part of the current European road network follows the old Roman roads. The *persistence of infrastructure* hypothesis is consistent with the main findings of the development economics literature studying the impact of the Romans' decisions.

The Roman institutions as a source of cultural convergence and trust

Different findings (Giavazzi et al., 2019; Desmet and Wacziarg, 2021; Jackson and Medvedev, 2024) emphasize a historically persistent cultural divide between and within countries. Cultural characteristics imply time, risk or social preferences that in turn determine economic behavior and linkages. These differences are likely to have deep historical roots (Becker et al., 2020) that may be attributable to some extent to political integration: Considering bilateral trust as a consequence of cultural convergence, Grosjean (2011) shows that part of Eastern European cultural heterogeneity is the result of the division between the Ottoman, Habsburg, Russian, and Prussian Empires. Her results suggest that social trust between communities becomes deeply impacted when political integration lasts more than four centuries. On the other hand, less persistent and more recent states, such as USSR and Yugoslavia, left social trust unchanged. In Poland, Grosfeld and Zhuravskaya (2015) detect regional differences in religious practices and in beliefs towards democratic ideals associated with the partition of Poland into 3 different empires during the 19th century. Nikolova and Plopeanu (2025) achieve similar conclusions associated with gender attitudes for Romania.

The Romans influenced Western European cultures: An early process of integration of the Italian peninsula into a relatively homogeneous Roman culture and identity between 350 and 100 BC is stated by the convergences of art and religion[12] while the Roman citizenship is progressively granted to communities in Northern Italy[13]. Following the conquest of a large western and eastern territory, the new Roman provinces received a similar treatment: their reorganization that started early under the reign of *Augustus* was based on multiple self-governing civic communities (*civitates*) belonging to the province, which were associated with

[12]See Bradley (2000) and Roselaar (2012), among others.
[13]See Cecchet et al. (2017) for a detailed discussion on Roman citizenship at that period.

a defined rural area politically subordinated to an urban capital. This urban center tended to reflect the republican image of Rome with a political duality of the *ordo decurionum* (local elites) and the *populus* (the rest of the represented population). These centrally devised imposed structures caused a relatively strong level of uniformity in both civic institutions and political procedures in the Western provinces (Mouritsen, 2014). The development of these structures, as well as the establishment of local cities similar to Northern Italian towns, is often considered as the major change arising from the Roman conquest[14]. The Roman law[15] also progressively developed to apply to the entire Empire population in 212 AC (Mousourakis, 2015, Chapter 7, pp. 240-243). These new elements may have resulted in cultural convergence as institutions and culture influence one another (Alesina and Giuliano, 2015). Before this generalization of Roman law, there was a continuous striving of local conquered communities to be granted the *ius Latii* (Latin rights of citizenship). These rights were gradually granted by successive emperors to local communities, mostly when they were considered politically and culturally advanced, as rewards for exceptional military assistance or due to the local presence of sizable Roman migrant communities or colonies. While it was rather exceptional under *Iulius Caesar* in Gaul, it became common under the Emperors[16] until Roman citizenship was granted to all free inhabitants by *Caracalla* in 212 AC. As a consequence, the subjugated communities were encouraged to imitate the Roman public and private social setting and create similar institutions. During the Empire era, some signs of convergence are the disappearance of local languages replaced by Latin (Clackson, 2014) as well as new Roman habits of consumption, or the use of Latin-rooted names (Beard, 2022, Chapter 12, pp. 616-623)[17]. The research of the (fiscally interesting) Latin rights and the other aspects of the political integration may have triggered cultural integration[18]. It

[14]See Millet (1990), Laurence and Trifilò (2015), pp. 108-10, and Beard (2022), Chapter XII, pp. 610-11.

[15]See section 1.B.3 for a detailed discussion on Roman law and its potential influence beyond cultural aspects.

[16]See Edmondson (2006) for a general review of this aspect of the Roman political integration.

[17]See also Woolf (2000) for a careful review of historical findings suggesting that the *Gallo-Roman* cultures were a mix of Roman and local elements, as well as Bowman and Wilson (2009), p.17, on the progressive apparition of typical Roman-consumed goods across the Empire.

[18]There were also evidence of local policies promoting the adoption of the Roman culture for the local

also created and/or encouraged the creation of similar political, cultural, and administrative institutions. It resulted in a partial cultural and linguistic convergence of the European regions that may have affected their economic linkages in the long run: most studies point to a positive impact of cultural and linguistic proximity on trade[19].

Roman-induced diasporas and migration patterns

In line with previously mentioned elements, Gomtsyan (2022) argues that the cities associated with past Roman infrastructure have higher current migration rates conditionally to local wealth, potentially because of a greater tolerance. The Roman institutions and infrastructure may be associated with specific attitudes toward migrants that facilitate migrations. It is also possible that Roman-rooted cultural convergence results in more migrations between European regions. These migration patterns may trigger regional trade thanks to *diaspora* networks[20].

The Roman integration as a source of Industrial divergences

The incorporation of a newly conquered territory into a relatively integrated market represents a major regional change. The trade was organized by the Roman state and private merchants simultaneously (Temin, 2012), implying a relatively liberal economy. According to differences in resources or technology, the change in transport costs that results from both political and economic integrations would then facilitate regional specializations to benefit from comparative advantages or new product varieties. Groot and Deschler-Erb (2015) indicate that regions in *Helvetia* (current Switzerland) typically started to produce chicken and pork meat at higher levels following the Roman integration, while provinces in the current

elites: for instance, *Tacitus* mentioned in *Agricola* that many sons of the Britain tribes elites received a subsidized Roman education. It was, however, relatively uncommon.

[19]See Felbermayr and Toubal (2010), Hellmanzik and Schmitz (2015), Tadesse and White (2010) or Guiso et al. (2009) among others for findings associated with a positive impact of cultural proximity, and Cyrus (2015) for evidence of an absence of effect. Additionally, Melitz and Toubal (2014) as well as Egger and Toubal (2016)'s findings support a positive impact of linguistic proximity beyond trust and ethnicity.

[20]See Genc et al. (2012) for a meta-analysis on the impact of *diaspora* and migrations on trade.

Netherlands specialized in cattle and horses. These early changes in specialization related to differences with new regional markets may have triggered regional-specific innovations and knowledge that, in turn, determine part of the industrial specializations in further periods.

1.3 Data

The bilateral trade intensity between 2011 and 2019 is retrieved from the European Road Freight Transport survey (ERFT) by aggregating the microdata on journeys and goods between the 127 Western NUTS2 regions. The exact location of departure and arrival of the trucks are not available, which means that NUTS2 regions are currently the most disaggregated level to study bilateral trade patterns in Europe. The ERFT vehicles' sample weights are applied to obtain representative estimates. To capture trade only, any journeys that are unladen or considered as a collection or distribution round, as well as goods that are not related to trade[21], are dropped following Santamaría et al. (2021). As the ERFT dataset provides freight weights, commodity prices by weight are estimated using the Trade Unit Values from the UN Comtrade database. It contains bilateral trade values and net weights at the 6-digit HS commodity level, allowing the construction of a (2011$) price index by commodity for each country pair. These indices are then aggregated at the NST 2007 industry[22] level using the Tan (2016)'s concordance table. It allows the computation of the total share of imports, trucks, and goods weights from an origin region in the sum total of a destination region. The ERFT data also provides information on the distance that the trucks traveled, allowing the estimation of a revealed trade distance in kilometers for each trading pair of regions. As additional trade and distances measurement robustness checks, I use the 2017 interregional input-output linkages and current trade costs expressed in (Euro) currency from the European Joint Research Centre[23]. These data provide an estimation of

[21] Mail, parcels, Equipment and material utilized in the transport of goods, Goods moved in the course of household and office removals, Grouped goods, Unidentifiable goods: goods which for any reason cannot be identified and therefore cannot be assigned to other groups, Other goods n.e.c.

[22] NST industries are listed in the section 1.B.8 table 1.B18.

[23] From García-Rodríguez et al. (2023) as well as Persyn et al. (2019).

the total cost of sending a truck between any (trading or not) regional pair of the sample, accounting for local policies and the network structure, but excluding regions belonging to Switzerland or Liechtenstein. They also provide geodesic distance between population-weighted regional centroids and input linkages proxies at the bilateral regional level.

Roman variables - Political and economic integration data under the Roman Empire are from Flückiger et al. (2022), who measured bilateral trade volumes during Roman imperial times using information on terra sigillata from the Roman tableware database. They compute the joint Roman rule century from the end of the Latin league; the political exposure of each region is depicted in fig. 1.2 (a), showing the patterns of the main independent variable. Unsurprisingly, countries participate differently in the variations due to the North-South pattern of the conquest. It is more visible in the section 1.B.1 that describes the bilateral source of variations: Spain, Portugal, and Italy form together the most strongly jointly integrated pairs, while pairs with at least one Northern country will have lower Roman integration values. This pattern requires careful attention to North-South patterns and Mediterranean access that may be an important confounding factor.

The multimodal Roman network at its full extension (117 AC), that is depicted in fig. 1.2 (b), allowed Flückiger et al. (2022) to compute the *Roman effective distance* as the average cost associated with shipping goods along the least cost path between two places, given the Roman transport network[24] and Roman-era-specific freight rates for each mode of transport[25] (with maritime trade cost as the reference). They also measure the 2018 cells' financial linkages as the share of firms in a destination grid cell j that are at least 25% owned by a firm in an origin grid cell i, using Orbis data between February and April.

Finally, other variables associated with past political integration of regions are computed from the *Cliopatra* dataset (Bennett et al., 2025) that contains time and spatial information

[24]They built this network combining elements from the Talbert and Bagnall (2000) digitalized version of the Barrington Atlas of the Greek and Roman World with various sources on navigable river sections and coastal paths, providing broader geographic coverage than the Stanford ORBIS project.

[25]Based on the Emperor Diocletian's Edict on Prices (301 AC).

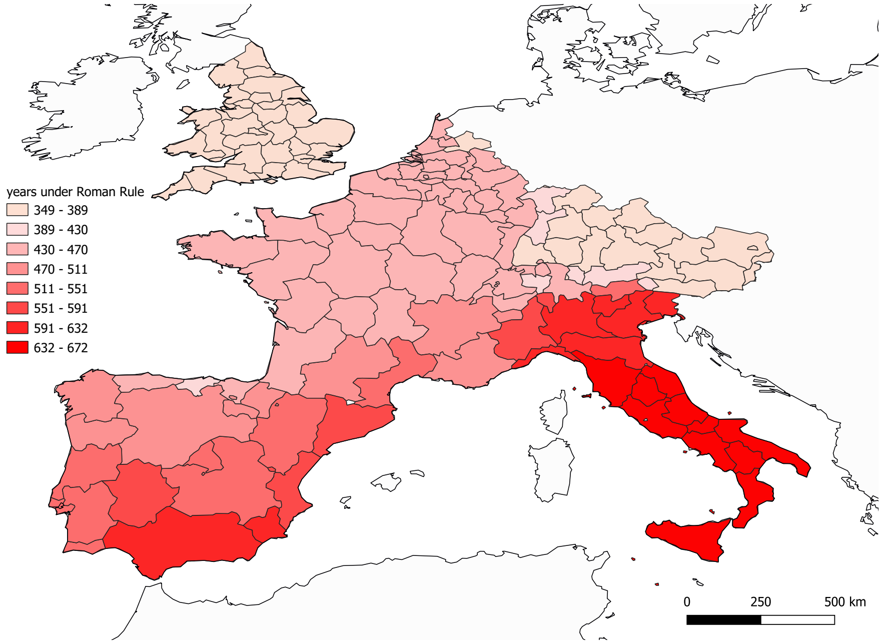
on most of the political entities that covered any parts of Europe between 91 BC and 2024 AC. It allows for the computation of regional past political exposure from any documented political entity. To do so, I only consider the head of the state as the state definition (For instance, I consider the Carolingian Empire instead of its subpart Kingdom of France). I aggregate these variables at the NUTS2 level following the same methodology based on geographic weights depicted in section 1.B.7. While the measure of joint Roman political integration is retrieved from Flückiger et al. (2022) as their measures go further than the first century BC, I complete the missing Eastern sample using these data for robustness checks.

Descriptive statistics of the bilateral NUTS 2 sample main variables are available in the table 1.1. A correlation matrix of the distances is available in the section 1.B.1 table 1.B1, as well as a broad descriptive analysis of the relationship between trade and transport costs in the section 1.B.1.

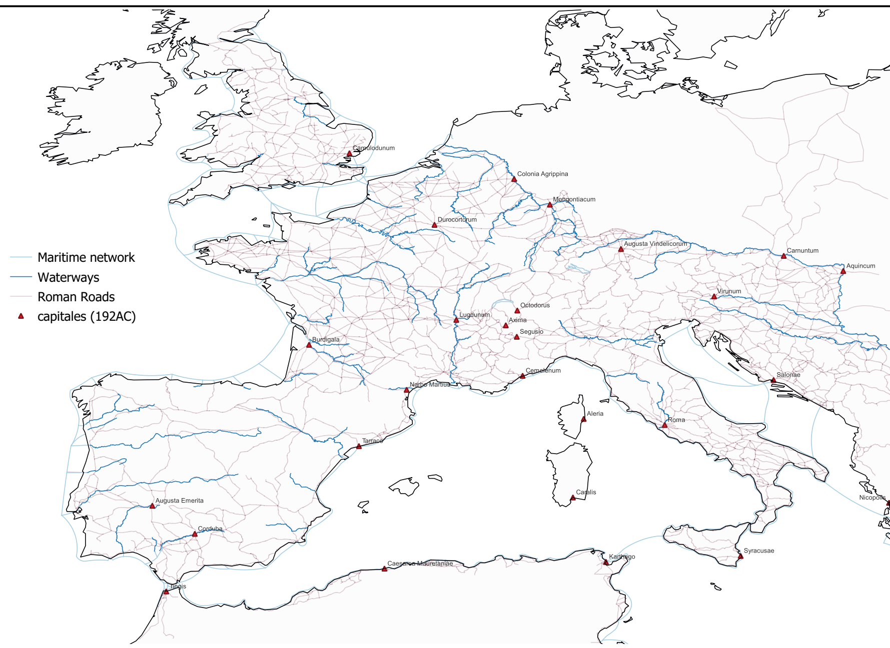
Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Trade share	0.0033	0.0149	0	0.6056	16,256
Weight share	0.0031	0.0133	0	0.4131	16,256
Transaction share (shipments)	0.0029	0.0129	0	0.4838	16,256
Inputs share	0.0037	0.0165	0	0.2794	14,042
Ownership share	0.0037	0.0204	0	0.6793	16,129
Light correlation	0.6834	0.1093	0.2241	0.9752	15,750
Joint Roman rule (cent.)	4.1141	0.6146	3.5700	6.7200	16,256
Tot. Joint Past rule	8.1754	4.9461	3.5700	22.9582	16,256
Effective Roman distance (log of)	9.4213	0.4412	6.9412	10.4643	16,256
Same State	0.1325	0.339	0	1	16,256
Av. shipment distance, km (log of)	6.6613	0.7847	2.7726	8.6622	12,668
Geodesic distance, km (log of)	6.5533	0.7574	0.4284	7.7475	16,256
Weight. geodesic distance, km (log of)	6.5715	0.7742	2.4942	7.7622	14,280
Transport costs, 2013 $ (log of)	7.2982	0.7875	3.9109	8.5500	14,280
Latitude difference	5.2627	3.8387	0.0201	18.783	16,256
Similar biome indicator	0.5274	0.4993	0	1	16,256
Waterway indicator	0.8784	0.3268	0	1	16,256
Mediterranean indicator	0.0311	0.1737	0	1	16,256

Note: Statistics computed on the Western NUTS2 bilateral sample pictured in section 1.B.8 fig. 1.B12 Source: Author's elaboration on Flückiger et al. (2022), *Cliopatra*, UN Comtrade, European Commission JRC and Eurostat ERFT data

Table 1.1: Summary statistics



(a)



(b)

Figure 1.2: 2013 NUTS2 European Regions Roman past political integration as the number of years under Roman rule (a) and 117AC Roman network (b)

1.4 Methodology

Main specification

This research relies on an identification advantage related to the early globalization shock that Roman integration represents: there were fewer economic links between the European economic units before their integration. In order to assess whether past common economic integration fosters later trade relationships between two geographical entities, I apply a standard gravity model (Anderson and Van Wincoop, 2003). The gravity equation for trade between geographical entities has the advantage of having various micro-foundations in the trade literature[26]. Applied to trade shares following Eaton et al. (2012) and separating trade costs into three different components, it is expressed as follows:

$$x_{ij} = exp(\omega Roman\ integration_{ij} + \delta ln(Roman\ Eff.\ Distance_{ij}) + \tau'_{ij}\theta + \gamma_i + \gamma_j + \epsilon_{ij}) \quad (1.1)$$

Where x_{ij} represents the bilateral trade share (i.e. the share of the imports of destination j from i in the total imports of j) or another measurement of common market integration. The main variables of interest are $Roman\ integration_{ij}$ which represents the Roman-era political common integration and is captured by the number of joint centuries the regions belonged to the Roman dominion, and the estimated effective distance cost under the Roman Empire $Roman\ Eff.\ distance_{ij}$ (economic integration). Standard errors are clustered at the regional pair level. As recommended by Silva and Tenreyro (2006) as well as Head and Mayer (2014), the model is estimated using PPML estimators to include the variations from the absence of trade between many European regions[27] and avoid the *heteroskedasticity* bias.

Empirically assessing the causal impact of past political integration on current trade is challenging: In this case, it requires that the Roman integration measurements are not associated with unobserved features, conditionally on control variables, but it is difficult to

[26] see Arkolakis et al. (2012) or Costinot and Rodríguez-Clare (2014) for a general review

[27] 21.13% of the 2011-2019 trade bilateral links are equal to 0 in the data while 36.8% had no financial linkages in 2018. See ppmlhdfe as a Stata command to deal with high-dimensional data.

discriminate between the Roman influence and other historical events and integration. For instance, an important share of modern roads is built on ancient Roman roads, as pictured by Dalgaard et al. (2022). Being able to understand whether roads and states originated from past Roman decisions or geographic connectivity and resource complementarity (the pre-trends challenge) is a first potential difficulty. To distinguish their influence from posterior events and integration (the legacy challenge) is also hardly feasible because numerous posterior states inherited their institutions both directly and indirectly. To distinguish the impact of past trade costs and integration from current ones and factors that would have resulted in later trade without the Romans' intervention, I control for current trade costs (measured by geodesic, fastest path network, and trucks revealed distances) as well as other bilateral common characteristics: belonging to the same current country, having a river, having similar biome, the latitude difference, and a Mediterranean indicator. τ'_{ij} contains these parts of the trade costs that are not directly related to the Roman era. Additionally, I replicate the instrumental strategy of Flückiger et al. (2022) who use a Gabriel graph of the Roman battles[28] as a fictional network that serves as an instrument of the Roman network; its variations are likely to capture military-oriented road creation rather than unobserved local features influencing the shape of the network. I also replicate their placebo strategy using pre-Roman trade intensity proxies to test for the presence of a pre-trend. Back to the posterior identification concerns, many subsequent states and institutions were inspired by the Roman law and structures, but it is hardly feasible to distinguish this influence from the direct integration associated with these states and institutions. To address these considerations, I control in a second step for the regional joint belongings to any known European state that existed in the current era. It helps to disentangle the Roman political exposure influence from other historical states and unobservable features facilitating political integration in general.

[28] A Gabriel graph is a geographic network built from a set of points (here Roman battles) based on the idea that points are linked if there is no third point closer to both of them. If a third point falls inside the diameter circle, then the direct connection between the two points is considered redundant and is removed (Matula and Sokal, 1980)

Another identification threat lies in the fact that a destination region may serve as a distribution center, artificially increasing its trade share with the other European regions. Secondly, some regions may have stronger ties with extra-European markets and have less importance as direct European trading partners. However, the origin and destination fixed effects γ_i and γ_j account for the related level of trade of the sending and the receiving regions (as well as their respective economic weights and multilateral resistance). To avoid capturing the potential remaining bilateral trade variations associated with a possible combination of these two factors (i.e. if one takes place at the origin and the other at the destination), I also drop regions with important international trade gateways in further regressions.

1.5 Main results

This section is structured as follows: Firstly, I present evidence of the influence of Roman political and economic integration on trade beyond infrastructure using gravity PPML estimates. To ensure that results are not driven by the way trade is measured, I also use alternative measurements of the early 21st-century and historical economic ties. Secondly, I briefly investigate the long-lasting influence of the Roman network on contemporary trade. Thirdly, I perform different sensitivity tests related to current political fragmentation, additional geographic bilateral features, and potential non-linearity in the relationship between the Roman political integration and trade linkages. Thirdly, I show that the main results are robust to the complete subsequent European political history. Fourthly, I apply a spatial Regression Discontinuity Design (RDD) exploiting the northern part of the Roman Limes as an identification strategy. Finally, I test several mechanisms that may explain the persistent Roman influence: cultural imprints and convergences, industrial specializations, migrations, and linguistic communities.

1.5.1 Duration of the Roman rule and later economic linkages

Baseline Estimates - The gravity equation (1.1) is estimated in the table 1.2 using the 21st-century freight trade share as the dependent variable[29]. The estimates can be interpreted as semi-elasticities of bilateral trade intensity to the duration of the joint Roman rule. All the specifications highlight a strong correlation between the Roman integration and current trade despite the covariates that capture bilateral geographic features (Mediterranean, waterways, and similarities in biome dummies as well as latitude differences that capture resource differences). The second and third columns indicate a decrease when geographic connectedness, measured by geodesic distance between regional centroids, and contemporary political integration are taken into account. The table 1.2 columns (4) and (5) show that it also survives the inclusion of current freight trade costs measured by the average distance of a shipment revealed by the data and the estimated total cost of sending a shipment between two regions[30]. In these specifications, where trade costs are the most appropriately measured with respect to the dependent variable, one more century of Roman integration implies that the concerned European regions' predicted trade intensity increases by 121.4%. This is considerable, especially because it is the influence of one single century of Roman integration. Adding a country x Same State fixed effect in the sixth column, capturing country-specific home bias in taste or differences in trade costs for foreign and domestic trade (Hillrichs and Vannoorenberghe, 2022), reduces this estimated semi-elasticity to 26.9%. While this effect remains sizable, this difference reflects the strong importance of the heterogeneity in the (contemporary) home bias/border effect captured in the estimate of the main specifications.

Alternative economic ties measures - Freight trade represents around 52% of the European

[29] See section 1.B.2 table 1.B2 for similar results with trade in level.

[30] The first can only be computed from trading pairs, while the second was not available for Switzerland and Liechtenstein, hence the decreases in the number of observations. The section 1.B.2 table 1.B4, that replicates columns (1) and (2) using the same observations as in the fifth column, shows that the decrease in the estimate is not solely driven by the exclusion of these two countries.

	(1) All	(2) All	(3) Trading pairs	(4) Trading pairs	(5) W. Switzerland	(6) W. Switzerland
Joint Roman rule (cent.)	3.260***	1.090***	1.036***	0.845***	0.795***	0.240*
	(0.272)	(0.165)	(0.156)	(0.153)	(0.123)	(0.139)
ln Roman eff. distance	-3.568***	-0.622**	-0.616**	-0.252	0.104	0.116
	(0.180)	(0.299)	(0.295)	(0.163)	(0.103)	(0.122)
Same state		1.406***	1.393***	1.354***	1.124***	
		(0.0804)	(0.0745)	(0.0686)	(0.0593)	
ln Geodesic distance		-1.436***	-1.423***	-0.593***	-0.515***	-0.865***
		(0.186)	(0.185)	(0.120)	(0.0730)	(0.111)
ln Shipment distance				-1.077***		
				(0.0979)		
ln Shipment cost					-1.540***	-1.152***
					(0.105)	(0.121)
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes
Geogr. controls τ'_{ij}	Yes	Yes	Yes	Yes	Yes	Yes
Country × Same state FE	No	No	No	No	No	Yes
Observations	16,256	16,256	12,667	12,667	14,280	14,280

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All models are estimated using Poisson Pseudo Maximum Likelihood (PPML). Robust standard errors clustered at the origin and destination NUTS2 regional levels are reported in parentheses. The dependent variable is total freight exports (2011$) from an origin to a destination NUTS2 region. *Joint Roman rule* measures the number of centuries during which both regions jointly belonged to the Roman Republic or Empire. *Roman effective distance* corresponds to the average least-cost transport path implied by the Roman transport network and Roman freight rates. *Geodesic distance* is the distance between regional centroids. *Same state* is a dummy equal to one when both regions belong to the same contemporary country. *Shipment distance* is the average truck shipment distance observed between 2011 and 2019 in ERFT data. *Shipment cost* measures the estimated cost of road shipment between regions. Geographic controls include latitude differences, Mediterranean access, waterways, and biome similarity indicators. All specifications include origin and destination fixed effects. Source: Author's calculations based on Flückiger et al. (2022), UN Comtrade, European Commission JRC, and Eurostat ERFT data.

Table 1.2: Gravity PPML estimates – Roman integration, transport connectivity, and bilateral trade intensity

trade within the continent according to Eurostat (Santamaria et al., 2023). To ensure the results presented above reflect broad economic linkages, the table 1.3 presents results using alternative measurements of European economic ties[31]. Additionally, as centroids may be relatively arbitrary tools to measure geographic distance, column (1) shows that the main results are unaffected when controlling for the (log of) geodesic distance between regional population-weighted centroids to account for the general location of human settlements and cities. The column (2) uses the trade share computed with direct weight rather than values in constant currency; the (in)significance of the estimates ensures the prices are not the only drivers of the previous findings in table 1.2 column (5). The column (3) uses the

[31]While the magnitudes vary, the estimates remain significant to the inclusion of a country-Same State fixed effect in section 1.B.2 table 1.B6.

input linkages intensity computed by the Joint Research Center as an alternative trade measurement. It allows for testing if the Roman integration influences the value chain spatial distribution; the estimates are highly similar to the coefficients obtained with trade shares. In column (4), the night-light correlation is a proxy for regional business cycle synchronization; the estimate is significant and positive. It is worth noting that it has the wrong expected sign for the Roman effective distance. The Roman political integration is also positively associated with financial linkage intensity, as depicted in column (5). The results seem remarkably stable across the different measures of economic linkages regarding the Roman political integration.

Intensive and extensive margins of trade - The regressions in table 1.2 and table 1.3 col. 5 capture both the intensive and the extensive margins of interregional trade and financial linkages. Decomposing the effects across the two margins in section 1.B.2 table 1.B3 reveals that the table 1.2 results are driven primarily by the intensive margin of trade, whereas both margins contribute to the observed effects for financial linkages. It may reflect the different nature of those ties as regional trade linkages are more common than financial linkages (See section 1.B.6).

Roman integration influence across time - Finally, if the Roman integration influence on economic linkages is causal, it should be persistent across time. To investigate this persistence, the analysis presented in the section 1.B.5 fig. 1.B9 (b) tests for a continuous influence across time of the Roman integration on trade from the Roman era to the early 21st century, showing gravity estimates based on column (3) specification (without the Same state control). It pictures a continuous positive influence of the Roman political integration on western economic linkages proxies, without an observable pre-trend.

	(1) Trade share	(2) Weight share	(3) Inputs share	(4) light corr.	(5) Ownership share
Joint Roman rule	0.867***	0.779***	0.739***	0.0405***	0.973***
	(0.130)	(0.135)	(0.127)	(0.0144)	(0.243)
ln Roman eff. distance	0.0735	-0.160	-0.0788	0.0357*	-0.716***
	(0.113)	(0.130)	(0.232)	(0.0200)	(0.255)
Same State	1.179***	1.204***	0.491***	-0.0116**	0.788***
	(0.0668)	(0.0642)	(0.0776)	(0.00589)	(0.158)
ln W. geodesic distance	-0.821***	-0.332**	1.426***	-0.103***	-1.930***
	(0.157)	(0.164)	(0.203)	(0.0206)	(0.341)
ln Shipment cost	-1.113***	-1.667***	-0.874***	0.0126	0.137
	(0.177)	(0.191)	(0.170)	(0.0177)	(0.335)
Origin FE	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes
Geogr. controls τ'_{ij}	Yes	Yes	Yes	Yes	Yes
Observations	14,280	14,280	14,042	13,806	13,924

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. The dependent variables are the share of (freight) exports from an origin to a destination NUTS2 region in the total imports of the destination region in values (1) and in weight (2), the night-time light intensity Pearsons correlation (3), the share of firms in destination that are at least > 25% owned by origin firms (4) and the share of inputs from origin region in the total import of inputs of the destination (5). The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. *geodesic distance* is the length (kilometers) between Regions computed using population-weighted centroids. *Same State* is a binary variable that captures common belonging to a country. All specifications include destination and origin Fixed effects and use PPML estimates (light corr. has also only positive values). Source: Author's computations on Flückiger et al. (2022), European Commission JRC and Eurostat ERFT data

Table 1.3: Gravity PPML estimates - Alternative economic ties measurements

1.5.2 Sensitivity analysis

This section performs a battery of additional sensitivity checks with respect to the results discussed above. It mostly consists of exploring the role of the *border effect*, the influence of geographic characteristics, nonlinearities, industrial heterogeneity, and sensitivity to the extension or restriction of the composition of the sample.

Border effect - The previous section notably emphasizes the importance of the heterogeneity of the border effect in the main baseline estimates. Digging further in this direction, the table 1.4 first two columns replicate the table 1.2 fourth column by splitting the sample between pairs belonging to the same contemporary state and different states. The estimates show that the joint Roman rule positive coefficients were mostly driven by international pairs; the coefficient for intranational pairs is insignificant and has the wrong sign. Adding interaction terms of the Roman-era integration variables in column (3) indeed shows that the current political integration attenuates the estimated effect of past integration. Taken together, these results suggest that the Roman integration has an attenuating influence on the *border effect*. It would imply that regions that are currently politically integrated may be relatively unaffected by past Roman integration, but that past integration seems to matter for trade between regions that are not politically integrated anymore.

Other geographic controls - The table 1.4 fourth column adds controls for differences in ruggedness, elevation, precipitation, and temperature; they do not affect the main results and are individually and jointly insignificant.

Allowing for nonlinearities - The fifth column tests for a non-linear effect of the joint Roman rule by using dummies of the century during which the two regions were integrated together for the first time into the same Roman state. The reference is the 2nd century BC. The results suggest a relatively linear effect of an additional joint century on (log of) bilateral trade intensity.

Industrial heterogeneity - One may worry that the main results are affected by a trade industrial aggregation bias (Redding and Weinstein, 2019). However, the industrial heterogeneity analysis performed in section 1.B.2 fig. 1.B4 reveals relatively homogeneous coefficients across the NST industries, comparable to the results presented in the previous section.

Spatial aggregation - Significant estimates in persistent studies can be related to biased standard errors due to spatial autocorrelation (Conley and Kelly, 2025). To address this concern, I replicate the baseline regression at the broader NUTS1 level while clustering the standard error at the bilateral NUTS1 level (section 1.B.6 fig. 1.B10 (b)). The estimate remains positive and significant[32].

Extension of the sample - Until now, the scope of the analysis was restricted to Western Europe for several reasons: Firstly, regional trade data are not available for North African and Middle-East regions. Preexisting networks before the Roman integration are also not always accurately documented, making it difficult to obtain results while conditioning for older infrastructure. Secondly, the Roman institutions were often mixed with preexisting local ones and were less homogeneous in this part of the Empire. Unsurprisingly, the extension of the sample to the Eastern (European) part of the Empire results in a substantial reduction of the estimate (table 1.B5), which may notably reflect a Greek pre-trend (Lomas, 2005; Creanza, 2024).

Robustness to resampling and other specifications - As additional robustness checks, the section 1.B.2 table 1.B7, table 1.B11, table 1.B12, and table 1.B13 show that the results are stable across various specifications and that no country solely drives the results (they survive the exclusion of any country from the sample).

Excluding main trade gateways - Finally, the trade shares may indirectly capture trade with regions outside Western Europe if goods produced elsewhere pass through the European main gateways and are then transported by trucks to their final destination. It may induce a positive bias in freight measurement for these specific regions. To avoid capturing extra-European trade, I therefore exclude the main European ports from the sample in the section 1.B.2 table 1.B9: The results remain unaffected.

[32]Replicating the NUTS2 regressions with standard errors clustered at the bilateral NUTS1 level leads to the same conclusion with a t-statistic equal to 3.50.

	(1) Same State	(2) Diff. States	(3) All	(4) All	(5) All
Joint Roman rule	-0.412	0.592***	1.218***	0.989***	
	(0.328)	(0.186)	(0.160)	(0.164)	
x Same State			-0.400***		
			(0.105)		
ln Roman eff. distance	-0.772**	-0.630***	-0.747**	-0.784***	-0.665**
	(0.347)	(0.200)	(0.298)	(0.258)	(0.299)
x Same State			0.254***		
			(0.0927)		
1st Century BC					-0.575***
					(0.153)
1st Century AC					-0.901***
					(0.207)
2nd Century AC					-1.596***
					(0.397)
Same state			0.939	1.323***	1.468***
			(0.908)	(0.0689)	(0.0916)
ln Geodesic distance	-1.228***	-1.711***	-1.457***	-1.191***	-1.432***
	(0.199)	(0.103)	(0.189)	(0.181)	(0.185)
Origin FE	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes
Geogr. controls τ'_{ij}	Yes	Yes	Yes	Yes	Yes
Other Geogr. controls	No	No	No	Yes	No
Observations	2,154	13,878	16,256	15,750	16,256

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. The dependent variable is the share of (freight) exports from an origin to a destination NUTS2 region in the total imports of the destination region in values. The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. *geodesic distance* is the length (kilometers) between Regional centroids. *Same State* is a binary variable that captures common belonging to a country. *x Same State* indicates an interaction of this variable with the variable above. All specifications include destination and origin fixed effects. Source: Author's own computations on Flückiger et al. (2022), European Commission JRC and Eurostat ERFT data

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.4: Gravity PPML estimates - sensitivity checks

1.5.3 Accounting for posterior political history

The previous findings emphasize the strong, positive, and continuous association between a common long-lasting political integration during the Roman era and trade intensity, as well as other economic ties. The Joint Roman Rule estimates proxy semi-elasticities of trade intensity to joint centuries under the Roman state. When computing exact variations, it appears that if two regions have one more century together under the Roman reign, their current predicted trade and financial intensities are multiplied by at least 2. Even if the means (exposed in table 1.1) of these variables are low[33], it would imply that the Roman Empire and Republic jointly have a predictive power for the 21st-century Western European market integration that is comparable to the current countries themselves. It may capture unrelated historical as well as current events, decisions, and features correlated with the Roman presence that affect economic linkages in the long run. Additionally, there may be natural determinants to the formation of states that affect the global political and economic history of Europe simultaneously. This section exploits the European subsequent political history to control for such unobservables and evaluate the importance of the Roman integration with respect to the broader European political history.

Comparison with global political integration - For the sake of comparison, the table 1.5 columns (1) and (2) PPML estimates allow to retrieve the elasticity of trade intensity to one more century of joint integration under any confounded state since 91 BC[34]. It barely reaches 1.132 with a coefficient approximately ten times lower than the equivalent for joint Roman rule in table 1.2, showing that the Roman influence is indeed considerable with respect to other historical integration. This difference reveals a strong heterogeneity: any past state does not have similar impacts, even if the net effect seems positive.

Accounting for political history - In the two following columns, I estimate the gravity (1.1) while controlling for each past political state's common exposure[35] to test if the trade impact of Roman political exposure survives. It appears that the estimated elasticity is almost divided by 2 when political history is taken into account, and shrinks to 1.24 once also controlling for more accurate proxies of transport costs than geodesic distance. Con-

[33]If the average destination region had one more century under Roman rule with the average origin region, the share of this region in its total imports would increase by approximately 0.5 pp. While it may seem negligible, it represents an increase of almost four standard deviations.

[34]The Total Joint past rule represents the global common duration of political integration. For instance, if two regions were unified for ten centuries in one state and then ten other centuries in another state, this variable would be equal to 20 for this pair, reflecting joint integration in any state.

[35]That is the equivalent of the Joint Roman rule variable for any documented European state. For instance, these covariates include the joint duration under the Holy Roman Empire, Prussia and the Napoleonic Empire.

trolling for posterior political history ensures the positive relationship detected previously is not solely driven by subsequent political integration and the natural determinants of state formation.

Accounting for a Catholic Church trend - However, European historical integration is not solely political, but also religious: the Catholic Church may have had a role in the early European economic integration, and its influence is likely to be correlated with the Roman integration[36]. I add a control in column (6) for the number of centuries of exposure to the Catholic church between the 6th and the 15th centuries[37]. While a positive and relatively strong association between the Catholic Church's influence during the Middle Ages and later trade intensity is indeed found (One more century two western regions have a (respective) bishop at the same time is associated with a 10% increase in their trade intensity), its inclusion does not negatively affect the magnitude and significance of the joint Roman rule estimate[38].

The Roman influence on subsequent trade linkages remains robust to the inclusion of alternative controls for historical integration. Unfortunately, these estimates are likely to underestimate the Roman integration influence on trade intensity: Most of the past states inherited the Roman institutions, and some, such as the Frankish kingdom, were even the result of the Romans' direct decisions. The fact that the coefficients remain sizable and significant despite such considerations implies that the Roman influence goes beyond these potential explanations, but these estimates require caution.

[36]The section 1.B.4 shows the influence of the Roman integration on the spread of Christian ideas.

[37]Joint Church exposure is computed from Schulz et al. (2019) who measured the Catholic Church influence at regional contemporary levels using the diffusion of bishoprics between 550 and 1500 AC.

[38]The Church influence could not be computed accurately for the entire sample, notably for the UK and Germany; this resampling is mostly responsible for the strong increase of the Roman rule estimate from table 1.5 fifth to sixth column: the estimate of specification (5) when excluding the missing regions is 0.418.

	(1) Trade share	(2) Trade share	(3) Trade share	(4) Trade share	(5) Trade share	(6) Trade share
Total Joint Past rule	0.124***	0.0436***				
	(0.0110)	(0.00706)				
Joint Roman rule			0.503***	0.218*	0.289**	0.475***
			(0.167)	(0.120)	(0.125)	(0.184)
Joint Church exposure						0.095*
						(0.049)
Same State		1.147***	1.415***	1.259***	1.298***	1.239***
		(0.0754)	(0.0818)	(0.0663)	(0.0715)	(0.072)
ln Geodesic distance	-2.007***		-2.087***	-0.734***		
	(0.0829)		(0.0795)	(0.104)		
ln W. geodesic distance		-0.694***			-0.749***	-0.916***
		(0.134)			(0.139)	(0.211)
ln Transport costs		-1.131***		-1.415***	-1.141***	-0.816***
		(0.163)		(0.121)	(0.185)	(0.219)
ln Roman eff. distance					0.273**	0.096
					(0.120)	(0.131)
Past states exposures	No	No	Yes	Yes	Yes	Yes
Geogr. controls τ'_{ij}	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,256	14,280	16,256	14,280	14,280	8,190

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. The dependent variable is the share of (freight) exports from an origin to a destination NUTS2 region in the total imports of the destination region. The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. *Joint Church exposure* is the number of centuries the regions were jointly under the influence of the Catholic Church from the 6th to the 15th centuries. It is not available at the NUTS2 level for the UK and DE. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. *geodesic distance* is the length (kilometers) between Regions computed using (population-weighted and geographic) centroids. *Same State* is a binary variable that captures common belonging to a country. All specifications include destination and origin Fixed effects and use PPML estimates. Source: Author's computations on Flückiger et al. (2022), Schulz et al. (2019), *Cliopatra*, European Commission JRC and Eurostat ERFT data

Table 1.5: Gravity PPML estimates - Past political exposure and trade intensity

1.5.4 Evidence from the Limes: Discontinuity regressions

The previous section attempts to approach causal identification of the Roman influence on contemporary trade by controlling for the whole political history of the European sample. It accounts for the idea that Western regions differ in resources or geographic characteristics that are key determinants of state formation, as well as for other political roots of economic linkages. However, it neglects the fact that European political history is partially a legacy of the Roman states' integration, and it may capture the influence of other important characteristics uncorrelated with subsequent political history. This section makes use of a spatial regression discontinuity approach[39] to answer these potential threats. It takes advantage of the Roman Limes that crosses Germany, the Netherlands, and the *Hadrian*'s wall in the United Kingdom as identification tools: Comparing neighboring regions that were jointly within or separated by the Roman Limes allows comparison between areas similar in unobserved characteristics affecting trade, as long as these did not determine the Roman Limes itself. The Roman Limes is defined using recent archeological estimates that are more accurate than traditional historical maps[40]. Any region (centroid) beyond this limit is considered non-Roman. A similar strategy was typically used by Wahl (2017) to assess the influence of Roman integration on wealth in current Germany. It relies on the assumption that unobserved elements affecting trade, such as geographic and climatic features, have continuous change at the Limes. This strategy can be adapted to bilateral trade. Consider the extensive Roman integration treatment in a gravity equation:

$$x_{ij} = exp(\rho_1 Roman_{ij} - \tau_{ij})\gamma_i \gamma_j \epsilon_{ij} \tag{1.2}$$

Where $Roman_{ij}$ is a treatment indicator based on (centroid) regional past joint belonging to the Roman Empire (2nd century AC). Assuming similar characteristics between two origin regions h and k that may differ in their side of the Limes and their respective iceberg transport costs with the destination region τ_{ij}, one can express the following trade ratio:

$$\frac{x_{hj}}{x_{kj}} = exp(\rho_1(Roman_{hj} - Roman_{kj}) - (\tau_{hj} - \tau_{kj}))\frac{\epsilon_{hj}}{\epsilon_{kj}} \tag{1.3}$$

If h and k are similar in any characteristics affecting trade that are not Roman-rooted but

[39]See Dell (2010) for a pioneering work adapting this research design to her study of the long-lasting impact of forced labor systems. See Egger and Lassmann (2015) for an early application to a gravity equation.

[40]The limes was retrieved by associating Dutch data on Limes sites from van Dinter (2011) with the UNESCO Limes definition (available here: https://whc.unesco.org/en/list/1631/maps/; note that the Limes does not reflect the maximum extent of the Roman Empire in the Netherlands.), adding the German shapefile provided by the German Limes Road Association (available here: https://www.limesstrasse.de/en/deutsche-limes-strasse/limes-road/gps-data-limes) that is consistent with the UNESCO definition. The Hadrian's wall is considered for the United Kingdom, using Collins and Symonds (2019)'s shapefile. The resulting border is pictured in fig. 1.3.

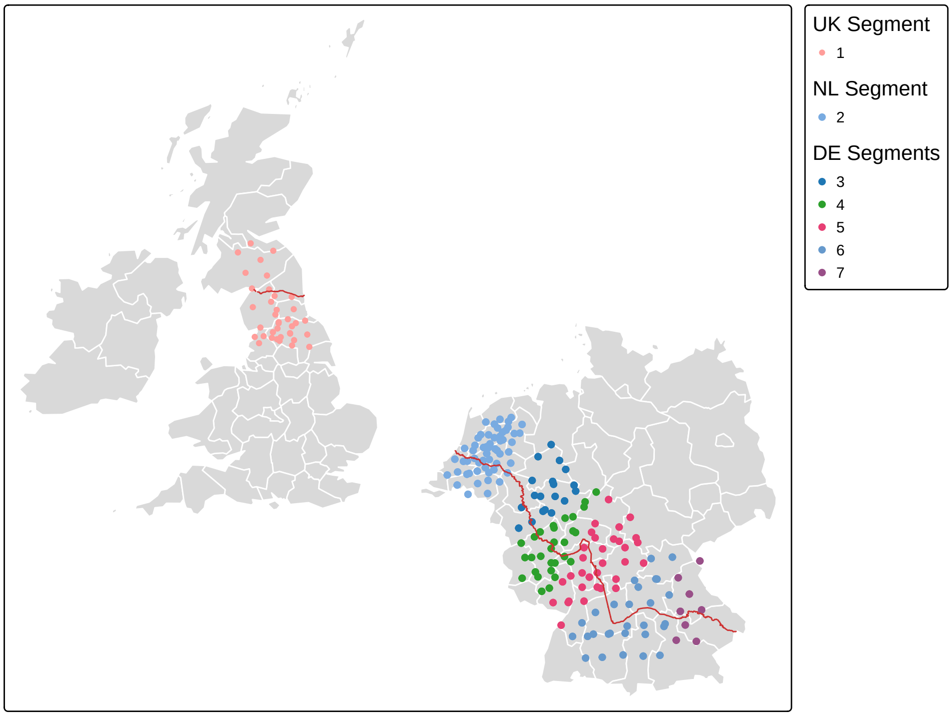
in observed characteristics τ, then independence between the error term and the treatment is achieved, and ρ_1, the causal impact of the integration of European territories in the Roman Empire on later trade, can be estimated[41]. To achieve such similarity, I limit the sample to German, Dutch, and UK regions located at 200km from each other and from the previously defined Limes. I also drop any international pair. Then, limiting the sample to the Roman-side destinations[42], ρ_1 can be estimated as a local average treatment effect (LATE) in the following spatial RDD specification:

$$x_{ij} = exp(\rho_0 + \rho_1 Roman_{ij} + \rho_2 D_i + \rho_3 D_i * Roman_{ij} + \gamma_{segment,ij} - \tau_{ij} + u_{ij}) \tag{1.4}$$

Where the forcing variable D_i^{Limes} is the (minimum) geodesic distance (km) of the origin region i's centroid to the Limes (multiplied by -1 when non-Roman). It allows for a sharp discontinuity regression design, as the treatment is mechanically assigned according to the side of the Roman Limes. The term τ_{ij} represents the trade costs between two regions, proxied by a linear combination of (log of) geodesic and revealed travel truck distance, as well as the shipment cost. The LATE ρ_1 can be interpreted as the difference in how a region at the Roman side of the Limes will trade (imports) more with a region located at the same side of the Roman border than with a similar region at the other side. Finally, the term $\gamma_{segment,ij}$ is a spatial segment fixed effect allowing for a within-segment estimation. Indeed, by using the distance to the Limes D_i^{Limes} as the forcing variable without taking accurate location into account, one ends up comparing pairs that are far apart from each other (Keele and Titiunik, 2015). To ensure proximity in the comparison, I split the Limes into seven equal segments and connect each pair of regions of the sample to the closest segment to their geographic midpoint, as depicted in fig. 1.3. The resulting segment fixed effect also captures

[41]It would, however, reflect a broad Roman influence, including Roman-rooted network connections and urban development.

[42]It creates the conditions for a sharp discontinuity design by assigning the treatment according to the origin region.



Note: The map depicts the German and Dutch Limes as well as Hadrian's wall (red lines) in the UK. Each dot represents the midpoint between the centroids of two NUTS2 regions that are located at less than 200km from each other and from the nearest Limes point. Only intranational pairs are represented. The Limes is segmented into seven equal parts to which each regional pair is assigned according to their midpoint and its nearest Limes part. Source: Author's construction based on UNESCO Limes definitions, as well as the German Limes Road Association and Eurostat shapefiles.

Figure 1.3: Roman Limes segments and regional bilateral midpoints

the country fixed effect. Finally, u_{ij} is an error term.

The estimates of eq. (1.4) in columns (1) and (2) indicate that a region that was part of the Roman Empire imports 50 to 55%[43] more from a nearby region that was also historically part of the Roman dominion than from a comparable nearby region located at the other side of the Limes. While this effect is sizable, it is lower than previous estimations as it describes a global effect of Roman integration rather than its intensive margins (over centuries). The estimates tend to increase when polynomial functions of the forcing variable are used, as in the quadratic case (columns (4) and (5)), but these additional variables are not jointly

[43]Exact variations. Respective p-values are 13.6% and 6.3%.

	(1)	(2)	(3)	(4)	(5)	(6)
	UK+NL+DE	DE	DE, no Rhine	UK+NL+DE	DE	DE, no Rhine
Roman	0.405	0.438*	0.261***	0.700**	0.402	0.535***
	(0.272)	(0.236)	(0.0779)	(0.308)	(0.329)	(0.159)
$f(D_i)$	Linear	Linear	Linear	Quadratic	Quadratic	Quadratic
τ_{ij} proxies	Yes	Yes	Yes	Yes	Yes	Yes
Segm. FE	Yes	Yes	No	Yes	Yes	No
N	202	116	36	202	116	36
ps. R^2	0.809	0.781	0.725	0.814	0.783	0.734

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, Note: robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. Columns (1) to (3) represent PPML estimates of eq. (1.4) while columns (4) to (6) add the squared distance to the limes and its treatment interaction among the covariates. Columns (1) and (4) use the whole sample of region pairs depicted in fig. 1.3. Other columns limit the sample to Germany or the part of Germany where the Limes is not following the Rhine. τ_{ij} controls include (log of) truck and geodesic population-weighted distances as well as the estimated trade cost. Source: Author's elaboration on Eurostat ERFT and JRC data.

Table 1.6: Spatial Regression Discontinuity Design estimates

significant. As Hadrian's Wall is correlated with the contemporary Scottish border, and the Limes follows the Rhine in the Netherlands and in its German Northern part, I limit the sample to regions located around the part of the German Limes seemingly uncorrelated with geographic features in columns (3) and (6). The main estimate shrinks to 30%. It becomes, however, highly local as it is built on a sample of 36 bilateral linkages out of 11 origin and 7 destination regions in South-West Germany. Overall, these results suggest a strong influence of Roman integration that may, however, also reflect Roman-induced development and access to an early large-scale trade network.

1.6 Roman integration and persistence: Mechanisms

In this section, I seek to test potential mechanisms related to the results presented in the previous sections. To do so, I apply the following specification using dependent variables that capture the potential mechanism described in section 1.2.2:

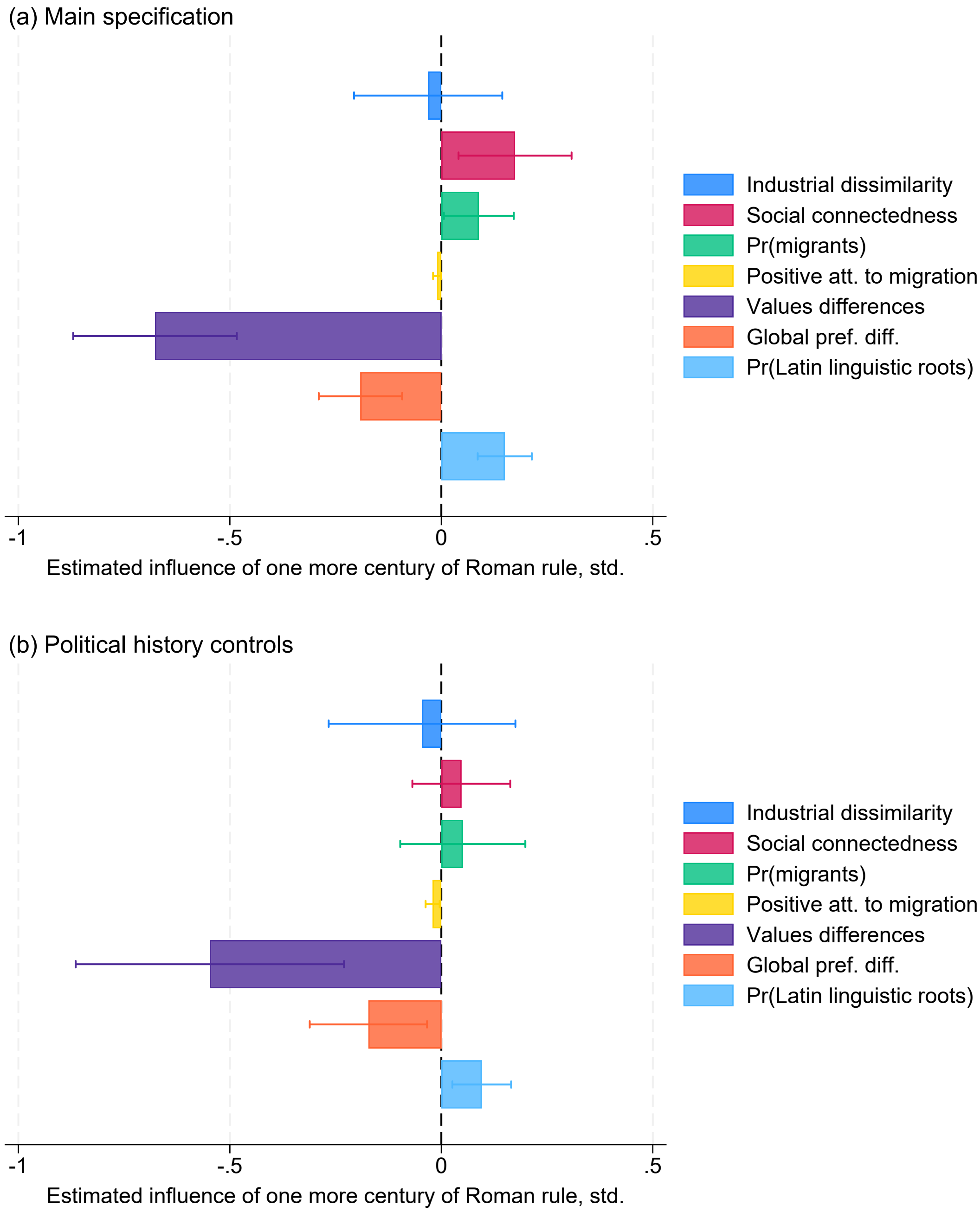
$$Mechanism_{ij} = \beta_1 Roman\, integration_{ij} + \beta_2 ln(Roman\, Eff.\, Distance_{ij}) + \tau_{ij}'\alpha + \gamma_i + \gamma_j + \epsilon_{ij} \quad (1.5)$$

The variable $Mechanism_{ij}$ is a standardized measure that captures a potential mechanism. The variable may either represent a difference or a similarity between regions i and j. τ'_{ij} contains the same control variables as in the baseline specification (1.1). This method allows for identifying correlates of the Roman political integration effect from the previous models. The main estimates using these dependent variables are represented in fig. 1.4 where the (a) specification is the (1.5) main model, and the (b) graph adds the political history controls. The dependent variables are detailed in each subsection and include migration, cultural, industrial, and institutional (dis)similarities measures. These equations are estimated using OLS estimates, as the standardized variables may have negative values.

1.6.1 Cultural and linguistic convergence

Cultural proximity - To test if the Roman influence is rooted in cultural convergence, I use three cultural distance or similarity measurements at the regional pair level as dependent variables. Exploiting new data from social media, I use the 2021 Facebook social connectedness index, first exploited by Bailey et al. (2018), that represents *Facebook (Meta)* ties between European regions, and I compute the absolute differences of the first dimension of a principal component analysis using the European Social Survey (ESS), capturing values or the Global Preferences (GPS) surveys. The first measure is likely to reflect both cultural proximity and bilateral migrations, while the two others capture differences in values, preferences, and perceptions. I find slight evidence of a Roman political exposure impact on social connectedness (fig. 1.4 red coefficients) that does not survive the inclusion of political history controls, while one more century under Roman rule is associated with a respective decrease of approximately 0.15 and 0.65 standard deviations in predicted Global preferences and Values absolute difference between the regions in the most strict specification (fig. 1.4 orange and purple coefficients). These results suggest a preference and cultural convergence consistent with the historical preference divide described by Becker et al. (2020).

Linguistic proximity - Additionally, the probability of drawing two persons speaking a



Note: The graphs display the joint centuries under Roman rule estimates and 95% confidence intervals (standard errors clustered at the origin and destination levels) from the eq. (1.5) specifications with various dependent variables. The τ'_{ij} vector includes biome similarity, waterways, Mediterranean, and same state dummies, as well as the (log of) weighted geodesic distances and freight trade cost as geographic connectivity control. Source: Author's computation on Flückiger et al. (2022), *Cliopatra*, Meta, ESS, GPS, and Eurobarometer data.

Figure 1.4: OLS estimates - Standardized mechanisms variables without (a) and with past states exposure controls (b)

Latin-rooted language separately in two regions is positively associated with the joint duration of Roman rule (light blue coefficients). It does not seem to be associated with the common presence of specific linguistic communities in two regions (section 1.B.2 fig. 1.B6 (a) shows no support for a significant link between the Roman rule and specific common languages in general, except for German if Austria and Germany are included in the sample), which means that individuals speaking the same Latin language do not drive this result. This element supports a long-run Latin cultural convergence.

Christian convergence - Finally, as the imperial period was characterized by the spread of Christianity across Europe, this detected cultural convergence may actually be driven by common past religious exposure rather than political integration. The section 1.B.4 attempts to test this potential channel. The table 1.B15 shows that more integrated regions were indeed more likely to have similar exposure to Christian communities, reflecting the past influence of the Roman network and political integration on the spread of new ideas. However, the table 1.B16 indicates that early Christian ideas and communities have no predictive power on contemporary trade.

Comparing Western and Eastern integration - As a placebo robustness check for cultural channels, I now exploit the Greek persistent culture during the Roman era. Different historical considerations make a long-lasting Roman direct influence more plausible in Western Europe than in the other parts of the Empire that experienced previous strong political, cultural, and economic integration, such as the Greek city-states, the Carthaginian state, or the Persian, Egyptian, and Babylonian Empires. The Latin cultural imprints in Greece were typically very limited (Woolf, 1994), while Greek cultural elements integrated the life of the Roman elites (Lomas, 2005; Beard, 2022, Chapter XII, pp. 620-622). As the Roman conquest had limited cultural impacts in this Eastern part of Europe, I exploit this difference between the Eastern and Western parts of the Empire by extending the dataset to the following countries: Albania, Bulgaria, Croatia, Greece, Hungary, Montenegro, Slovenia, and Romania. Comparing this part of the Empire, more exposed to Greek culture, to the

Western part, where Latin convergence is documented, can serve as a placebo regression for investigating cultural transmission channels. The section 1.B.2 table 1.B5 indeed pictures insignificant estimates for this area, consistent with a cultural convergence channel.

1.6.2 Migration and tolerance

To investigate the tolerance channel emphasized by Gomtsyan (2022), I compute the interaction of 1) the regional migrant shares (That is, the probability of drawing two migrants in the respective populations) and 2) the shares of people with positive attitude toward migrations (interpreted in the same way). The first measure will capture the diasporas channel in its indirect way (for instance, Belgians in France trading with Belgians in Belgium will not appear in these measures, but well if they trade with Belgians in the Netherlands) and a general joint exposure to migrations. While a positive influence on the predicted probability of drawing two migrants appears in fig. 1.4 (a), it does not survive the political history controls in (b), and there is no evidence of an influence on positive attitudes toward migrants (green coefficients). Applying the same approach to citizenship shares in section 1.B.2 fig. 1.B6 (b), I test for the citizenship-specific European diasporas' channel and find no evidence of remaining variations from such European networks, even from the German communities, despite the positive correlation with the German speakers.

Investigating the idea of Roman-rooted tolerance in a broader way, I test if the regions that experienced a longer Roman state exposure together are both more tolerant in general, or have more trust in people and institutions. To do so, I use the European Social Survey and voting data compiled by Algan et al. (2017). The coefficients are depicted in the section 1.B.2 fig. 1.B5. It provides no significant estimates for the interaction of voting shares for anti-establishment parties, nor for general trust in people, in the legal system, and in the European institutions.

1.6.3 Industrial divergences

Finally, I use the industrial dissimilarity index used by Flückiger et al. (2022) to test the specializations channel (Blue coefficients in fig. 1.4). The coefficient in (a) and (b) have the wrong expected sign and are insignificant. Using this variable, I find no support for long-run industrial specializations associated with the Roman rule.

1.7 Discussion

This chapter documents that the European political history influences long-run economic linkages between regions. More particularly, it emphasizes the strong influence on early 21st-century freight trade and financial ties of the longest and widest integration Europe experienced: the Roman republican and imperial integration. Its estimated influence is associated with the heterogeneity in the contemporary *Border effect*: The regions belonging to different current states that were strongly integrated during the Roman era tend to have stronger linkages, while the effect seems more limited for regions currently integrated. As a consequence, the Roman rule appears as an attenuating factor in the commercial negative impact of Western Europe's current fragmentation into different states. Even within the same state, comparing similar German regions from each part of the Roman Limes integration using a spatial RDD identification strategy leads to a significant Roman trade effect: a region on the Roman side trades 30% to 55% more with a neighboring region on the same side than with a neighboring region on the German side.

The major part of this influence is associated with geographic features and road infrastructure, as well as subsequent political integrations that partially have Roman roots. No evidence of an influence of human networks proxied by migrants' diasporas, nor of industrial specializations, is found; instead, the results point to cultural convergence as a major root of this persistence: The Roman conquest is associated with the Latin-rooted languages as well as with similarities in preferences or values. Taken together, these results support that Roman political unification fostered durable cultural and economic linkages that have continued to shape the spatial structure of Western European trade over two millennia.

1.A Bibliography

Alesina, A., Giuliano, P., 2015. Culture and institutions. Journal of economic literature 53, 898–944.

Algan, Y., Guriev, S., Papaioannou, E., Passari, E., 2017. The European trust crisis and the rise of populism. Brookings papers on economic activity 2017, 309–400.

Anderson, J.E., Van Wincoop, E., 2003. Gravity with gravitas: A solution to the border puzzle. American economic review 93, 170–192.

Arkolakis, C., Costinot, A., Rodríguez-Clare, A., 2012. New trade models, same old gains? American Economic Review 102, 94–130.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., Wong, A., 2018. Social connectedness: Measurement, determinants, and effects. Journal of Economic Perspectives 32, 259–280.

Beard, M., 2022. S.P.Q.R. : Histoire de l'ancienne Rome. Perrin.

Becker, A., Enke, B., Falk, A., 2020. Ancient origins of the global variation in economic preferences, in: AEA Papers and Proceedings, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203. pp. 319–323.

Becker, S.O., Boeckh, K., Hainz, C., Woessmann, L., 2016. The empire is dead, long live the empire! long-run persistence of trust and corruption in the bureaucracy. The Economic Journal 126, 40–74.

Bekker-Nielsen, T., 2006. Rome and the Black Sea Region: Domination, Romanisation, Resistance. volume 5. Aarhus Universitetsforlag.

Bennett, J.S., Mutch, E., Tollefson, A., Chalstrey, E., Benam, M., Cioni, E., Reddish, J., Zsambok, J., Levine, J., Cook, C.J., et al., 2025. Cliopatria - A geospatial database of world-wide political entities from 3400BCE to 2024CE. Scientific Data 12, 247.

Blaydes, L., Paik, C., 2021. Trade and political fragmentation on the silk roads: The economic effects of historical exchange between China and the muslim east. American Journal of Political Science 65, 115–132.

Boehm, J., Chaney, T., 2024. Trade and the end of antiquity. Centre for Economic Performance Discussion Paper .

Botsas, E.N., 1987. Greece and the east: The trade connection, 1851–1984. Journal of Modern Greek Studies 5, 207–235.

Bowman, A., Wilson, A., 2009. Quantifying the Roman economy: integration, growth, decline? Quantifying the Roman economy: methods and problems 1, 1.

Bradley, G., 2000. Ancient Umbria: state, culture, and identity in central Italy from the Iron Age to the Augustan era. Oxford University Press.

Cecchet, L., et al., 2017. greek and Roman citizenship: State of research and open questions, in: Citizens in the Graeco-Roman World. Aspects of Citizenship from the Archaic Period to AD 212.. Brill, pp. 1–30.

Chaney, T., 2014. The network structure of international trade. American economic review 104, 3600–3634.

Christakos, G., Olea, R.A., Serre, M.L., Wang, L.L., Yu, H.L., 2005. Interdisciplinary public health reasoning and epidemic modelling: the case of black death. Springer.

Clackson, J., 2014. Local languages in Italy and the West, in: Bruun, C., Edmondson, J. (Eds.), The Oxford Handbook of Roman Epigraphy. Oxford Academic. chapter 32, p. 699–720.

Collins, R., Symonds, M., 2019. Hadrian's wall 2009-2019. A Summary of Excavation and Research Prepared for the Fourteenth Pilgrimage of Hadrian's Wall .

Conley, T.G., Kelly, M., 2025. The standard errors of persistence. Journal of International Economics 153, 104027.

Costinot, A., Rodríguez-Clare, A., 2014. Trade theory with numbers: Quantifying the consequences of globalization, in: Handbook of international economics. Elsevier. volume 4, pp. 197–261.

Coughlin, C.C., Novy, D., 2021. Estimating border effects: The impact of spatial aggregation. International Economic Review 62, 1453–1487.

Creanza, P.P., 2024. Institutions, trade, and growth: The ancient greek case of Proxenia. The Journal of Economic History 84, 1–39.

Cyrus, T.L., 2015. Culture and trade in the European union. Journal of Economic Integration , 206–239.

Dalgaard, C.J., Kaarsen, N., Olsson, O., Selaya, P., 2022. Roman roads to prosperity: Persistence and non-persistence of public infrastructure. Journal of Comparative Economics 50, 896–916.

De Benedictis, L., Licio, V., Pinna, A.M., 2023. From the historical Roman road network to modern infrastructure in Italy. Journal of Regional Science 63, 1162–1191.

De Luca, G., 2016. Infrastructure financing in medieval Europe, in: Cassis, Y., De Luca, G., Florio, M. (Eds.), Infrastructure Finance in Europe: Insights into the History of Water, Transport, and Telecommunications. Oxford University Press. chapter 1, pp. 39–60.

Delabastita, V., Hoste, J., Maes, S., 2026. Political fragmentation and the spatial allocation of human capital .

Dell, M., 2010. The persistent effects of Peru's mining mita. Econometrica 78, 1863–1903.

Desmet, K., Wacziarg, R., 2021. The cultural divide. The Economic Journal 131, 2058–2088.

Dincecco, M., Onorato, M.G., 2016. Military conflict and the rise of urban Europe. Journal of Economic Growth 21, 259–282.

van Dinter, M., 2011. Landscape reconstruction of the western part of the Limes-zone in the Netherlands, V3. URL: https://doi.org/10.17026/DANS-2A4-TRW4, doi:10.17026/DANS-2A4-TRW4.

Disdier, A.C., Mayer, T., 2007. Je t'aime, moi non plus: Bilateral opinions and international trade. European Journal of Political Economy 23, 1140–1159.

Djankov, S., Freund, C., 2002. Trade flows in the former Soviet Union, 1987 to 1996. Journal of Comparative Economics 30, 76–90.

Drake, H.A., 2002. Constantine and the Bishops: the Politics of Intolerance. JHU Press.

Eaton, J., Kortum, S.S., Sotelo, S., 2012. International trade: Linking micro and macro. Technical Report. National bureau of economic research.

Edmondson, J., 2006. Cities and Urban Life in the Western Provinces of the Roman Empire 30 BCE-250 CE. A companion to the Roman Empire , 250–280.

Egger, P.H., Lassmann, A., 2015. The causal impact of common native language on international trade: Evidence from a spatial regression discontinuity design. The Economic Journal 125, 699–745.

Egger, P.H., Toubal, F., 2016. Common spoken languages and international trade, in: The Palgrave handbook of economics and language. Springer, pp. 263–289.

Evans, C.L., 2003. The economic significance of national border effects. American Economic Review 93, 1291–1312.

Federico, G., Schulze, M.S., Volckart, O., 2021. European goods market integration in the very long run: from the Black Death to the First World War. The Journal of Economic History 81, 276–308.

Feenstra, R.C., 2002. Border effects and the gravity equation: Consistent methods for estimation. Scottish Journal of Political Economy 49, 491–506.

Felbermayr, G.J., Toubal, F., 2010. Cultural proximity and trade. European Economic Review 54, 279–293.

Flückiger, M., Hornung, E., Larch, M., Ludwig, M., Mees, A., 2022. Roman transport network connectivity and economic integration. The Review of Economic Studies 89, 774–810.

Fousek, J., Kaše, V., Mertel, A., Vŷtvarová, E., Chalupa, A., 2018. Spatial constraints on the diffusion of religious innovations: The case of early christianity in the Roman empire. PLoS One 13, e0208744.

Frensch, R., Fidrmuc, J., Rindler, M., 2023. Topography, borders, and trade across Europe. Journal of Comparative Economics .

Fritsch, M., Obschonka, M., Wahl, F., Wyrwich, M., 2023. On the Roman origins of entrepreneurship and innovation in germany. Regional Studies , 1–18.

García-Rodríguez, A., Lazarou, N., Mandras, G., Salotti, S., Thissen, M., Kalvelagen, E., 2023. A NUTS-2 European Union interregional system of Social Accounting Matrices for the year 2017: The RHOMOLO V4 dataset. Technical Report. JRC Working Papers on Territorial Modelling and Analysis.

Genc, M., Gheasi, M., Nijkamp, P., Poot, J., 2012. The impact of immigration on international trade: A meta-analysis, in: Migration impact assessment. Edward Elgar Publishing, pp. 301–337.

Giavazzi, F., Petkov, I., Schiantarelli, F., 2019. Culture: Persistence and evolution. Journal of Economic Growth 24, 117–154.

Gomtsyan, D., 2022. The Roman origins of modern migration .

Groot, M., Deschler-Erb, S., 2015. Market strategies in the Roman provinces: different animal husbandry systems explored by a comparative regional approach. Journal of Archaeological Science: Reports 4, 447–460.

Grosfeld, I., Zhuravskaya, E., 2015. Cultural vs. economic legacies of empires: Evidence from the partition of Poland. Journal of Comparative Economics 43, 55–75.

Grosjean, P., 2011. The weight of history on European cultural integration: A gravity approach. American Economic Review 101, 504–08. URL: https://www.aeaweb.org/articles?id=10.1257/aer.101.3.504, doi:10.1257/aer.101.3.504.

Guiso, L., Sapienza, P., Zingales, L., 2009. Cultural biases in economic exchange? The quarterly journal of economics 124, 1095–1131.

Havranek, T., Irsova, Z., 2017. Do borders really slash trade? A meta-analysis. IMF Economic Review 65, 365–396.

Head, K., Mayer, T., 2000. Non-Europe: the magnitude and causes of market fragmentation in the EU. Review of World Economics 136, 284–314.

Head, K., Mayer, T., 2013. What separates us? Sources of resistance to globalization. Canadian Journal of Economics/Revue canadienne d'économique 46, 1196–1231.

Head, K., Mayer, T., 2014. Gravity equations: Workhorse, toolkit, and cookbook, in: Handbook of international economics. Elsevier. volume 4, pp. 131–195.

Head, K., Mayer, T., Ries, J., 2010. The erosion of colonial trade linkages after independence. Journal of international Economics 81, 1–14.

Hellmanzik, C., Schmitz, M., 2015. Virtual proximity and audiovisual services trade. European Economic Review 77, 82–101.

Hillberry, R., Hummels, D., 2008. Trade responses to geographic frictions: A decomposition using micro-data. European Economic Review 52, 527–550.

Hillrichs, D., Vannoorenberghe, G., 2022. Trade costs, home bias and the unequal gains from trade. Journal of International Economics 139, 103684.

Hopkins, K., 1980. Taxes and trade in the Roman empire (200 bc–ad 400). The Journal of Roman Studies 70, 101–125.

Jackson, J.C., Medvedev, D., 2024. Worldwide divergence of values. Nature Communications 15, 2650.

Jones, E.L., 2003. The European miracle: environments, economies and geopolitics in the history of Europe and Asia. Cambridge University Press.

Keele, L.J., Titiunik, R., 2015. Geographic boundaries as regression discontinuities. Political Analysis 23, 127–155.

Kraus, C., et al., 2014. Tacitus: Agricola. Cambridge University Press.

Lankhuizen, M., De Groot, H.L., Linders, G.J.M., 2011. The trade-off between foreign direct investments and exports: The role of multiple dimensions of distance. The world economy 34, 1395–1416.

Laurence, R., Trifilò, F., 2015. The global and the local in the Roman empire: Connectivity and mobility from an urban perspective, in: Globalisation and the Roman world: World history, connectivity and material culture. Cambridge University Press Cambridge, pp. 99–122.

Lomas, K., 2005. Rome and the Western Greeks, 350 BC-AD 200: conquest and acculturation in Southern Italy. Routledge.

Marczinek, M., Maurer, S., Rauch, F., 2025. Networks in trade-evidence from the legacy of the hanseatic league. Journal of International Economics , 104102.

Matula, D.W., Sokal, R.R., 1980. Properties of Gabriel graphs relevant to geographic variation research and the clustering of points in the plane. Geographical analysis 12, 205–222.

Mazhikeyev, A., Edwards, T.H., 2021. Post-colonial trade between Russia and former Soviet republics: back to big brother? Economic Change and Restructuring 54, 877–918.

McCallum, J., 1995. National borders matter: Canada-US regional trade patterns. The American economic review 85, 615–623.

Mees, A.W., 2018. Was there a difference between Roman civil and military Samian (terra sigillata) market supply? Finding answers with statistical distribution analysis methods. Internet Archaeology 50.

Melitz, J., Toubal, F., 2014. Native language, spoken language, translation and trade. Journal of International Economics 93, 351–363.

Millet, M., 1990. The Romanization of Britain. Cambridge University Press.

Mokyr, J., 2016. A culture of growth: The origins of the modern economy. Princeton University Press.

Mouritsen, H., 2014. Local elites in Italy and the western provinces, in: Bruun, C., Edmondson, J. (Eds.), The Oxford Handbook of Roman Epigraphy. Oxford Academic. chapter 12, p. 227–249.

Mousourakis, G., 2015. Roman law and the origins of the civil law tradition. Springer.

Nikolova, E., Plopeanu, A., 2025. Empires, gender attitudes and tolerance: Evidence from Romania .

Obschonka, M., Wahl, F., Fritsch, M., Wyrwich, M., Rentfrow, P.J., Potter, J., Gosling, S.D., 2025. Roma eterna? Roman rule explains regional well-being divides in germany. Current Research in Ecological and Social Psychology 8, 100214.

Ortman, S.G., Lobo, J., Lodwick, L., Wiseman, R., Bulik, O., Harbison, V., Bettencourt, L.M., 2024. Identification and measurement of intensive economic growth in a Roman imperial province. Science Advances 10, eadk5517.

Pascali, L., 2017. The wind of change: Maritime technology, trade, and economic development. American Economic Review 107, 2821–2854.

Persson, K.G., Sharp, P., 2015. An economic history of Europe. Cambridge University Press.

Persyn, D., Díaz-Lanchas, J., Barbero, J., 2019. Estimating road transport costs between EU regions. JRC working papers on territorial modelling and analysis .

Pétrequin, P., Cassen, S., Errera, M., Klassen, L., Sheridan, A., Pétrequin, A.M., 2012. Jade: grandes haches alpines du Néolithique européen, Ve et IVe millénaires av. J.-C. Centre de Recherche Archéologique de la Vallée de l'Ain.

Price, S., 2012. Religious mobility in the Roman empire. The Journal of Roman Studies 102, 1–19.

Redding, S.J., Weinstein, D.E., 2019. Aggregation and the gravity equation. AEA Papers and Proceedings 109, 450–455.

Roselaar, S.T., 2012. Processes of integration and identity formation in the Roman Republic. Brill.

Santamaria, M., Ventura, J., Yeşilbayraktar, U., 2023. Exploring European regional trade. Journal of International Economics 146, 103747.

Santamaría, M.A., Ventura, J., Yeşilbayraktar, U., 2021. Borders within Europe. Technical Report. National Bureau of Economic Research.

Schauer, P., Bevan, A., Shennan, S., Edinborough, K., Kerig, T., Pearson, M.P., 2020. British neolithic axehead distributions and their implications. Journal of Archaeological Method and Theory 27, 836–859.

Schulz, J.F., Bahrami-Rad, D., Beauchamp, J.P., Henrich, J., 2019. The church, intensive kinship, and global psychological variation. Science 366, eaau5141.

Silva, J.S., Tenreyro, S., 2006. The log of gravity. The Review of Economics and statistics 88, 641–658.

Tadesse, B., White, R., 2010. Cultural distance as a determinant of bilateral trade flows: do immigrants counter the effect of cultural differences? Applied Economics Letters 17, 147–152.

Talbert, R.J., Bagnall, R.S., 2000. Barrington Atlas of the greek and Roman world. Princeton University Press.

Tan, S.W., 2016. The Geography of Trade in the European Single Market. Technical Report. The World Bank.

Temin, P., 2012. The Roman market economy. Princeton University Press.

Wahl, F., 2017. Does European development have Roman roots? evidence from the German limes. Journal of Economic Growth 22, 313–349.

Ward, M.D., Hoff, P.D., 2007. Persistent patterns of international commerce. Journal of Peace Research 44, 157–175.

Wilson, A., Bowman, A., 2017. Trade, commerce, and the state in the Roman world. Oxford University Press.

Woolf, G., 1994. Becoming Roman, staying greek: culture, identity and the civilizing process in the Roman east. The Cambridge Classical Journal 40, 116–143.

Woolf, G., 2000. Becoming Roman: the origins of provincial civilization in Gaul. Cambridge University Press.

Yon, J.B., 2004. The Romanization of palmyra and the cities on the euphrates, in: Annales. Histoire, Sciences Sociales, Editions de l'EHESS. pp. 313–336.

1.B Appendix

1.B.1 Descriptive analysis appendix

This section presents complementary descriptive elements and a broad correlation analysis of the main variables used in the analysis.

Bilateral Roman integration: sources of variations

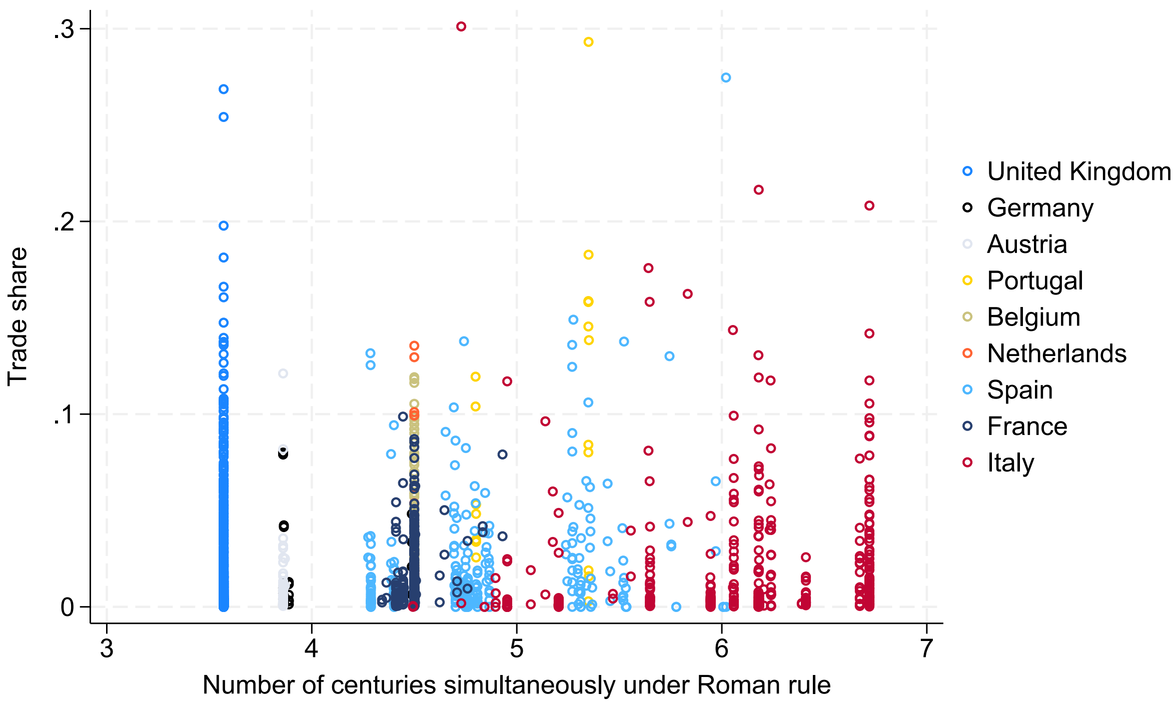
As there are 16,256 regional pairs in the main sample, picturing the variations behind the variables is complex. The scatterplot in fig. 1.B1 pictures the general variations in the dependent and independent variables from the intranational pair, which are easier to depict and are also informative on the international variations behind the main estimates. Due to heavy waves of Roman conquests, England, Germany, Belgium, Austria, and the Netherlands do not bring much intranational information, but they contribute importantly when paired with regions from Southern countries. Spain, Italy, Portugal, and, to a lesser extent, France bring intranational variations and represent the most integrated regions.

Distances correlation matrix

The table 1.B1 displays the correlation matrix of the main covariates used in the regression analysis. Interestingly, the period of common belonging to the Roman State is almost uncorrelated to most of the geographic distances, including Roman effective distance. Unsurprisingly, current geographic distance measurements are highly correlated with each other.

Variables	Roman rule	Roman dist.	ship. dist.	Geod. dist.	w. geod. dist.	Transp. costs
J. Roman rule	1.000					
Roman eff. distance	-0.016	1.000				
Av. shipment distance	-0.016	0.454	1.000			
Geodesic distance	-0.070	0.406	0.949	1.000		
Weight. Geodesic distance	-0.023	0.459	0.958	0.995	1.000	
Transport costs	-0.068	0.456	0.944	0.973	0.973	1.000
Latitude difference	-0.154	0.137	0.706	0.731	0.722	0.710

Table 1.B1: Distances cross-correlations



Note: Each point represents a national pair of regions. As a consequence, the most left points are associated with regions that have the lowest Roman political exposure, and *vice-versa* for the right. The dispersion gives an approximate idea of the variations within and between countries. Source: Author's computation on Eurostat ERFT, UN Comtrade, and Flückiger et al. (2022) data.

Figure 1.B1: Scatter plot of regional pairs trade share and Roman political joint exposure, within country pairs only

Financial and commercial linkages

The section 1.5.1 emphasizes different results between most of the economic ties measurements and the financial ownership shares used by Flückiger et al. (2022); the Roman network connectivity is indeed positively associated with this measurement of financial ties, while it is insignificant and/or is associated with a positive estimate when controlling for current trade costs. In this subsection, I investigate the potential reasons behind this difference. Using a Kernel regression to depict the general relationship between the February-April 2018 ownership shares and the 2011-2019 trade shares, one obtains the fig. 1.B2 picture. The non-parametrically estimated relationship has a relatively linear form, suggesting that the two variables indeed capture similar economic ties to a certain threshold. It is less evident when looking at regions highly integrated financially or commercially, which may explain a Pearsons correlation (0.57) lower than expected. Another major difference lies in their extensive margins: 22% of observations are associated with 0 trade, while 37.48% have no financial ties.

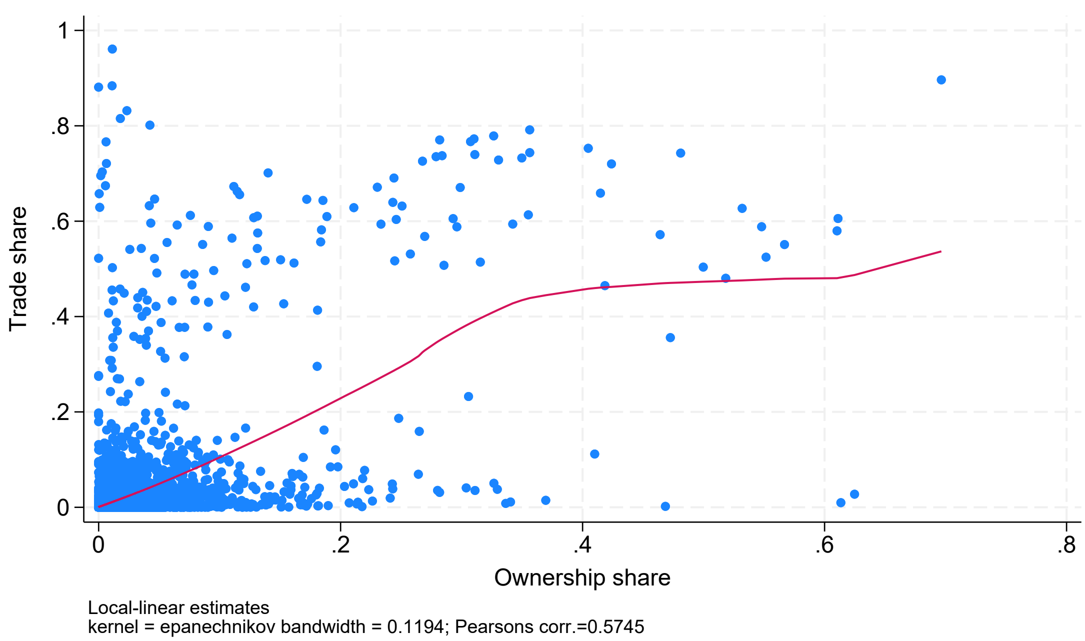


Figure 1.B2: Kernel regression - trade and ownership shares

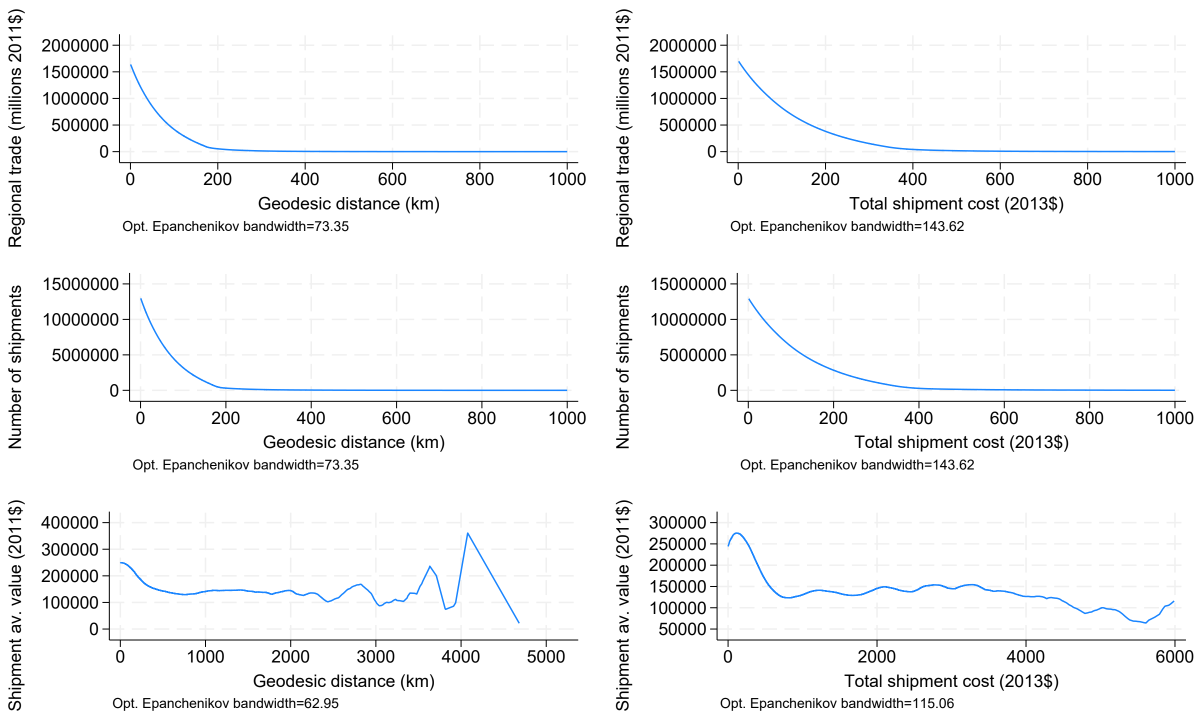
The relationship between European freight trade and transport costs

The top right Kernel regression in fig. 1.B3 draws a general relationship between regional bilateral freight trade and trade costs for the average estimated cost of a shipment: the trade strongly decreases with the trade cost in Europe, becoming flat at small levels once the estimated cost of sending a truck exceeds 450$ (2013 ref.). The top left graph pictures the relationship with geodesic distance: The shape is relatively similar to Hillberry and Hummels (2008), who realised this exercise on the US regional freight trade; their results showed a flat relationship when exceeding 320km *versus* 425km here, but their research level is a smaller zip regions pair instead of administrative units[44]. The total trade flows T_{ij} (from an origin region i to destination j) is the product of the total number of shipments N_{ij} and the average value of a shipment $\overline{P_{ij}}\ \overline{Q_{ij}}$:

$$T_{ij} = N_{ij}\ \overline{P_{ij}}\ \overline{Q_{ij}} \tag{1.B1}$$

Where $\overline{P_{ij}}$ and $\overline{Q_{ij}}$ are respectively the average good price and the average goods quantity of a shipment. When decomposing the previous relationship between its extensive N_{ij} and its intensive $\overline{P_{ij}}\ \overline{Q_{ij}}$ roots in the other graphs, it appears European freight trade is mostly shaped by its extensive margins: Europeans do not seem to compensate higher trade costs by sending more valuable goods but they rather send less shipments. The results suggest that the most valuable cargoes are sent to close destinations; however, this does not drive alone the main relationship between trade and transport costs.

[44]The geodesic distance is here computed using population-weighted centroids, but it remains possible that the difference is associated with the fact that geodesic distances are higher with wider areas because of their internal component: The centroids do not necessarily reflect most accurately the final destinations.



Note: Each graph depicts the relationship between trade components levels and costs from a Kernel regression performed using optimal Epanchenikov bandwidth. The left column presents results associated with geodesic distance, and the right column with the total estimated cost of sending a truck from origin to destination as the explanatory variables. The first line is associated with the total 2011-2019 freight trade between the two region T_{ij}, the second the total number of shipments N_{ij} and the third with the average value of a shipment $\overline{P_{ij}}\ \overline{Q_{ij}}$. The sample is the EU member states. Source: Author's computation on Eurostat ERFT and European Commission JRC datasets.

Figure 1.B3: Kernel regressions - trade and transport costs

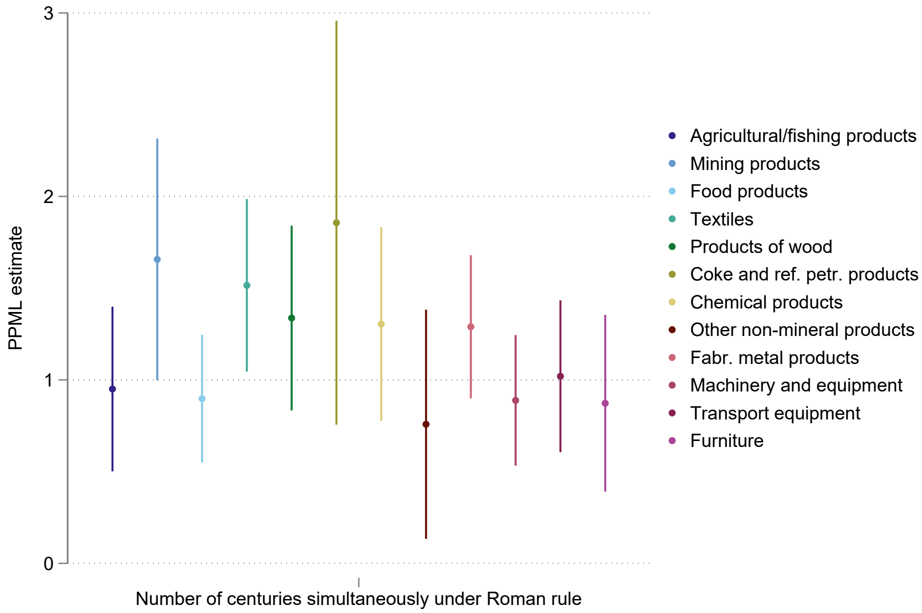
1.B.2 Regression analysis appendix

	(1) All	(2) All	(3) All	(4) Trading pairs	(5) Trading pairs	(6) W. Switzerland
Joint Roman rule	5.619***	4.210***	1.429***	1.420***	1.113***	0.969***
	(0.424)	(0.436)	(0.218)	(0.213)	(0.257)	(0.193)
ln Roman eff. distance	-4.489***	-3.428***	0.345**	0.345**	0.577***	0.773***
	(0.374)	(0.399)	(0.168)	(0.169)	(0.127)	(0.148)
Same State			1.533***	1.515***	1.590***	1.092***
			(0.103)	(0.100)	(0.114)	(0.0700)
ln Geodesic distance			-1.902***	-1.902***	-0.631***	-0.913***
			(0.0832)	(0.0827)	(0.109)	(0.158)
ln Shipment distance					-1.337***	
					(0.0936)	
ln Shipment cost						-1.534***
						(0.168)
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes
Geogr. controls τ'_{ij}	No	Yes	Yes	Yes	Yes	Yes
Observations	16,256	16,256	16,256	12,671	12,671	14,280

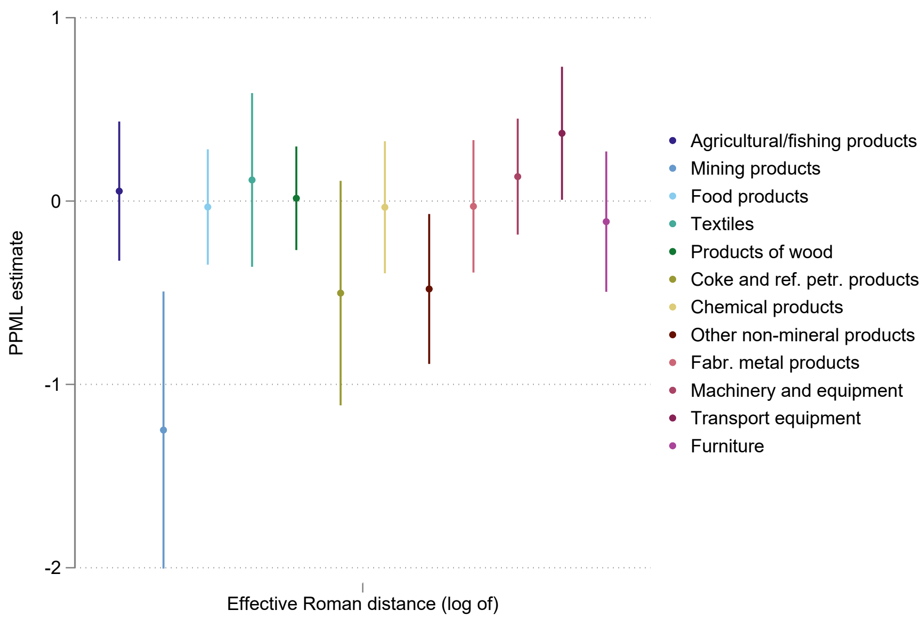
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regional level are reported in parentheses. The dependent variable is the share of (freight) exports from an origin to a destination NUTS2 region in the total imports of the destination region. The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. *Geodesic distance* is the length (kilometers) between two Regions' centroids. *Same State* is a binary variable that captures common belonging to a (current) country. *Shipment distance* is the average distance (km) that trading trucks made between the two regions between 2011 and 2019 as revealed by ERFT data (available for trading pairs only). *Shipment costs* is the estimated price of sending a truck from one region to another (available for non-trading pairs, but not for Switzerland and Liechtenstein). Geographic controls include latitude difference, as well as Mediterranean, waterways, and similar biome indicators. All specifications include destination and origin fixed effects. Source: Author's elaboration on Flückiger et al. (2022), UN Comtrade, European Commission JRC and Eurostat ERFT data

Table 1.B2: Gravity PPML estimates - Duration of joint Roman rule, Roman effective distance and bilateral regional trade (level)



(a)



(b)

Note: Based on the replications of the section 1.5.1 table 1.2 column (3) model, this graph depicts the PPML estimates and associated 95% conf. interval computing the trade shares for each 2007 NST industry subject to trade (section 1.B.8 table 1.B18). Source: Author's computation on Eurostat ERFT, Flückiger et al. (2022), and European Commission JRC datasets.

Figure 1.B4: Trade share and Roman integration, PPML estimates - Results by NST 2007 industry, Joint Roman rule (a) and Roman effective distance (b)

	(1) Trade	(2) # ownerships	(3) Trade share	(4) Ownership share	(5) Pr(Trade)	(6) Pr(Ownership)
Joint Roman rule	1.060***	0.723**	0.778***	0.571**	0.0883	0.616***
	(0.257)	(0.284)	(0.122)	(0.264)	(0.0563)	(0.133)
ln Roman eff. distance	0.408*	-1.116***	0.0433	-0.750***	-0.100*	-0.195*
	(0.220)	(0.308)	(0.110)	(0.244)	(0.0529)	(0.105)
Same State	1.162***	1.033***	1.159***	0.706***	-0.00244	0.357***
	(0.0872)	(0.170)	(0.0630)	(0.145)	(0.0253)	(0.0519)
ln W. geodesic distance	-0.968***	-2.023***	-0.738***	-1.894***	0.0965	-0.254*
	(0.242)	(0.369)	(0.154)	(0.331)	(0.0753)	(0.130)
ln Shipment costs	-1.173***	0.466	-1.207***	0.111	-0.192**	0.297**
	(0.306)	(0.362)	(0.171)	(0.316)	(0.0761)	(0.136)
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes
Geogr. controls τ'_{ij}	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14,280	13,924	11,391	8,724	14,280	13,924

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regional level are reported in parentheses. The dependent variable is (1) total trade, (2) the total number of ownerships, (3) the trade share among trading pairs only, (4) the ownership share among linked regions only, (5) the indicator of trade and (6) the indicator of ownsership. The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. *geodesic distance* is the length (kilometers) between Regions computed as a bilateral cells average. textitSame State is a binary variable that captures common belonging to a country. All specifications include destination and origin fixed effects. Source: Author's replication on Flückiger et al. (2022), UN Comtrade, European Commission JRC and Eurostat ERFT data

Table 1.B3: Gravity PPML estimates - Extensive and intensive margins

	(1) W. Switzerland	(2) W. Switzerland	(3) W. Switzerland	(4) Trading pairs	(5) Trading pairs	(6) W. Switzerland
Joint Roman rule	2.074***	2.095***	1.103***	1.044***	0.932***	0.867***
	(0.192)	(0.196)	(0.159)	(0.151)	(0.148)	(0.130)
ln Roman eff. dist.	-0.0199	-0.0837	0.0170	-0.0101	0.131	0.0735
	(0.141)	(0.146)	(0.116)	(0.115)	(0.115)	(0.113)
ln W. geodesic dist.	-1.913***	-1.839***	-1.684***	-1.672***	-0.787***	-0.821***
	(0.0534)	(0.0600)	(0.0532)	(0.0544)	(0.0847)	(0.157)
Same state			1.309***	1.297***	1.262***	1.179***
			(0.0628)	(0.0596)	(0.0586)	(0.0668)
ln Shipment dist.					-0.977***	
					(0.0862)	
ln Shipment cost						-1.113***
						(0.177)
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes
Geogr. controls τ'_{ij}	No	Yes	Yes	Yes	Yes	Yes
Observations	14,280	14,280	14,280	11,391	11,391	14,280

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regional level are reported in parentheses. The controls include the population-weighted bilateral geodesic distance, unavailable for Switzerland. The dependent variable is the share of (freight) exports from an origin to a destination NUTS2 region in the total imports of the destination region. The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. *Same State* is a binary variable that captures common belonging to a (current) country. *Shipment distance* is the average distance (km) that trading trucks made between the two regions between 2011 and 2019 as revealed by ERFT data (available for trading pairs only). *Shipment costs* is the estimated price of sending a truck from one region to another (available for non-trading pairs, but not for Switzerland and Liechtenstein). Geographic controls include latitude difference, as well as Mediterranean, waterways, and similar biome indicators. All specifications include destination and origin fixed effects. Source: Author's replication on Flückiger et al. (2022), UN Comtrade, European Commission JRC and Eurostat ERFT data

Table 1.B4: Gravity PPML estimates - Roman integration and regional trade intensity: Excluding Switzerland and Liechtenstein

	(1) Western	(2) Eastern	(3) Both	(4) Western	(5) Eastern	(6) Both
Joint Roman rule (cent.)	0.871***	0.095	0.100**	0.460***	-0.428	0.0787
	(0.123)	(0.314)	(0.049)	(0.134)	(0.356)	(0.0632)
ln Geodesic distance	-1.825***	-1.975***	-1.852***	-1.738***	-1.069**	-1.682***
	(0.0395)	(0.212)	(0.0373)	(0.0724)	(0.432)	(0.0634)
Mediterranean ind.	-0.0863	1.750***	0.505***	0.124	-1.254	0.210
	(0.128)	(0.607)	(0.175)	(0.133)	(0.820)	(0.149)
Same state	1.329***	1.874***	1.517***	1.247***	3.041***	1.224***
	(0.0651)	(0.493)	(0.0671)	(0.0786)	(0.980)	(0.0758)
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes
Political history	No	No	No	Yes	Yes	Yes
Observations	18,906	1,370	39,638	18,906	1,354	39,626

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regional level are reported in parentheses. The western sample is the main sample, while the Eastern sample contains the European (without Turkey) eastern part of the Roman Empire, as well as the Mediterranean islands. The dependent variable is the share of (freight) exports from an origin to a destination NUTS2 region in the total imports of the destination region. The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. *Same State* is a binary variable that captures common belonging to a (current) country. All specifications include destination and origin fixed effects. Political history controls contain the joint duration under the rule of each common era state. Source: Author's replication on Flückiger et al. (2022), UN Comtrade, *Cliopatra*, European Commission JRC and Eurostat ERFT data

Table 1.B5: Gravity PPML estimates - Roman integration and regional trade intensity, Western and Eastern European samples

	(1) Trade share	(2) Weight share	(3) Inputs share	(4) Nightlight corr.	(5) Ownership share
Joint Roman rule (cent.)	0.229*	0.290*	0.765***	0.0665***	0.601***
	(0.135)	(0.164)	(0.193)	(0.0142)	(0.201)
ln Roman eff. distance	-0.0623	-0.257*	-0.540**	0.00618	-0.716***
	(0.122)	(0.142)	(0.239)	(0.0167)	(0.236)
ln W. geodesic distance	-1.102***	-1.873***	1.144***	-0.138***	-1.873***
	(0.160)	(0.395)	(0.331)	(0.0277)	(0.395)
ln Shipment cost	-0.681***	0.0163	-0.484	0.0626**	0.0163
	(0.193)	(0.414)	(0.375)	(0.0275)	(0.414)
Origin FE	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes
Geogr. controls τ'_{ij}	Yes	Yes	Yes	Yes	Yes
Country x Same State FE	Yes	Yes	Yes	Yes	Yes
Observations	14,280	14,280	14,042	13,806	13,924

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level in parentheses. The dependent variable are the share of (freight) exports from an origin to a destination NUTS2 region in the total imports of the destination region in values (1) and in weight (2), the night-time light intensity Pearsons correlation (It has only positive values) (3), the share of firms in destination that are at least > 25% owned by origin firms (4) and the share of inputs from origin region in the total import of inputs of the destination (5). All specifications include destination and origin as well as Same State x Country Fixed effects. Source: Author's computations on Flückiger et al. (2022), European Commission JRC and Eurostat ERFT data

Table 1.B6: Gravity PPML estimates - Alternative economic ties measurements with country-specific border fixed effect

	(1) Own. Share	(2) Own. Share	(3) Own. Share	(4) Own. Share	(5) Own. Share	(6) Own. Share
ln Roman eff. distance	-5.613***	-1.034***	-0.838***	-0.838***	-0.680***	-0.818***
	(0.353)	(0.239)	(0.243)	(0.243)	(0.220)	(0.259)
Joint Roman rule (cent.)		1.034***	1.256***	1.256***	0.900***	0.583**
		(0.279)	(0.336)	(0.336)	(0.246)	(0.256)
ln Geodesic distance		-1.928***	-2.311***	-2.311***	-1.234***	-1.222***
		(0.0893)	(0.158)	(0.158)	(0.174)	(0.207)
Same State			0.423***	0.423***	0.500***	0.609***
			(0.157)	(0.157)	(0.140)	(0.153)
ln Shipment distance					-0.996***	
					(0.103)	
ln Shipment cost						-0.987***
						(0.236)
Observations	15,876	15,876	15,876	15,876	12,606	13,924

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. The dependent variable is the share of firms located in an origin region that are (> 25%) owned by firms located in a destination region. The sample is restricted to trading pairs in column 5 and excludes Liechtenstein and Switzerland in column 6. All specifications include destination and origin fixed effects as well as latitude difference + common country belonging, biome similarity, waterways, and Mediterranean dummies. Source: Author's replication on Flückiger et al. (2022), European Commission JRC and Eurostat ERFT data

Table 1.B7: Gravity PPML estimates - Financial linkages and Roman effective distance, a replication of Flückiger et al. (2022)

	(1) Trade Share	(2) Trade Share	(3) Trade Share	(4) Trade Share	(5) Trade Share	(6) Trade Share
ln Roman eff. distance	-6.611***	-0.812**	-1.302***	-1.018***	0.0228	0.0309
	(0.343)	(0.400)	(0.350)	(0.380)	(0.245)	(0.246)
Joint Roman rule		2.129***	1.051***	1.127***	0.798***	0.873***
		(0.265)	(0.183)	(0.191)	(0.117)	(0.122)
ln Geodesic distance		-1.694***	-1.212***	-1.267***	-0.496***	
		(0.186)	(0.164)	(0.176)	(0.0697)	
Same State			1.397***	1.329***	1.124***	1.175***
			(0.0796)	(0.0763)	(0.0575)	(0.0641)
ln Shipment cost					-1.519***	-1.123***
					(0.114)	(0.172)
ln W. geodesic distance						-0.788***
						(0.171)
Observations	15,500	15,500	15,500	13,572	13,572	13,572

Note: All the models are estimated using the instrumental Poisson Pseudo Maximum Likelihood (IV PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. The dependent variable is the share of trade from an origin to a destination NUTS2 region in the total imports of the destination region. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. It is instrumented using distances from a Gabriel graph[a] based on the Roman battles, from Flückiger et al. (2022). The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. *geodesic distance* is the length (kilometers) between Regions computed using population-weighted centroids. *Same State* is a binary variable that captures common belonging to a country. All specifications include destination and origin fixed effects and use instrumental PPML estimates. Source: Author's computations on Flückiger et al. (2022), European Commission JRC and Eurostat ERFT data.

Table 1.B8: IV PPML estimates - Trade linkages and Roman effective distance

[a]A Gabriel graph is a geographic network built from a set of points (here Roman battles) based on the idea that points are linked if there is no third point closer to both of them. If a third point falls inside the diameter circle, then the direct connection between the two points is considered redundant and is removed (Matula and Sokal, 1980)

	(1)	(2)	(3)	(4)	(5)	(6)
	Trade share	Trade share	Inputs share	Inputs share	Ownership share	Ownership share
Joint Roman rule (cent.)	0.734***	0.646***	0.810***	0.684***	1.431***	1.092***
	(0.173)	(0.147)	(0.161)	(0.164)	(0.384)	(0.254)
ln Roman eff. distance	-0.594***	-0.0959	-0.0318	-0.256	-0.953***	-0.891***
	(0.176)	(0.129)	(0.307)	(0.340)	(0.281)	(0.227)
Same State	1.365***	1.144***	0.603***	0.384***	0.361**	0.829***
	(0.0925)	(0.0771)	(0.133)	(0.124)	(0.163)	(0.169)
ln Geodesic distance	-1.909***		0.579***		-2.437***	
	(0.0919)		(0.122)		(0.197)	
ln w. geodesic distance		-0.492***		1.418***		-2.196***
		(0.169)		(0.271)		(0.401)
ln Shipment cost		-1.477***		-0.900***		0.366
		(0.212)		(0.214)		(0.406)
Observations	12,544	10,816	10,712	10,712	12,321	10,609

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. The dependent variable are either the share of (freight) exports from an origin to a destination NUTS2 region in the total imports of the destination region in values, the share of firms in destination that are at least $> 25\%$ owned by origin firms and the share of inputs from origin region in the total import of inputs of the destination. The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. The specifications also include among the covariates the common country belonging, biome similarity, waterways, and Mediterranean dummies as well as the latitude difference. All specifications include destination and origin Fixed effects and use PPML estimates (light corr. has also only positive values). Source: Author's computations on Flückiger et al. (2022), European Commission JRC and Eurostat ERFT data

Table 1.B9: Gravity PPML estimates - Excluding regions with main trade gateways

	(1) Trade Share	(2) Trade Share	(3) Trade Share	(4) Trade Share
Joint Roman rule	1.154***	0.934***	1.103***	0.867***
	(0.174)	(0.152)	(0.159)	(0.130)
ln Roman eff. distance	-0.489	0.0406	0.0170	0.0735
	(0.312)	(0.128)	(0.116)	(0.113)
Same State	1.332***	1.255***	1.309***	1.179***
	(0.0766)	(0.0589)	(0.0628)	(0.0668)
ln Geodesic distance	-1.439***	-0.505***		
	(0.189)	(0.106)		
LNtruckdist		-1.229***		
		(0.0880)		
ln W. geodesic distance			-1.684***	-0.821***
			(0.0532)	(0.157)
ln Shipment cost				-1.113***
				(0.177)
Observations	14,520	11,450	14,280	14,280

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. The sample excludes any pair including a region from the current Switzerland or Liechtenstein. The dependent variables is the share of (freight) exports from an origin to a destination NUTS2 region in the total imports of the destination region in values. The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. The specifications also include among the covariates the common country belonging, biome similarity, waterways, and Mediterranean dummies, as well as the latitude difference. All specifications include destination and origin fixed effects and use PPML estimates. Source: Author's computations on Flückiger et al. (2022), European Commission JRC and Eurostat ERFT data

Table 1.B10: Gravity PPML estimates - Excluding Switzerland

	(1) Italy	(2) France	(3) Spain	(4) Portugal	(5) Belgium	(6) Netherlands	(7) UK	(8) Germany	(9) Austria
Joint Roman rule	1.348***	0.524***	0.814***	0.844***	0.635***	0.850***	0.876***	0.851***	0.985***
	(0.216)	(0.130)	(0.163)	(0.140)	(0.127)	(0.138)	(0.117)	(0.134)	(0.129)
ln Roman eff. distance	0.134	0.0732	0.0817	0.0589	0.0326	0.0645	-0.199	0.0910	0.117
	(0.117)	(0.129)	(0.119)	(0.114)	(0.133)	(0.123)	(0.125)	(0.122)	(0.116)
Same State	1.122***	1.169***	1.176***	1.174***	1.327***	1.180***	1.186***	1.196***	1.181***
	(0.0676)	(0.0859)	(0.0824)	(0.0748)	(0.0734)	(0.0753)	(0.0605)	(0.0735)	(0.0689)
ln W. geodesic distance	-0.901***	-0.809***	-0.750***	-0.771***	-0.961***	-0.850***	-0.757***	-0.843***	-0.865***
	(0.165)	(0.191)	(0.177)	(0.160)	(0.149)	(0.155)	(0.215)	(0.160)	(0.155)
ln Shipment cost	-1.048***	-1.123***	-1.256***	-1.186***	-0.993***	-1.081***	-0.933***	-1.080***	-1.057***
	(0.194)	(0.231)	(0.203)	(0.178)	(0.181)	(0.181)	(0.222)	(0.180)	(0.177)
Observations	9,900	9,702	10,712	13,110	11,772	13,340	8,010	12,882	12,882

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. The sample excludes any pair including a region from the current country indicated in the column name. The dependent variable is the share of (freight) exports from an origin to a destination NUTS2 region in the total imports of the destination region in values, the share of firms in the destination that are at least $> 25\%$ owned by origin firms and the share of inputs from the origin region in the total import of inputs of the destination. The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. The specifications also include among the covariates the common country belonging, biome similarity, waterways, and Mediterranean dummies, as well as the latitude difference. All specifications include destination and origin fixed effects and use PPML estimates. Source: Author's computations on Flückiger et al. (2022), European Commission JRC, UN Comtrade and Eurostat ERFT data

Table 1.B11: Gravity PPML estimates - Excluding each country individually

	(1) Trade share	(2) Weight share	(3) light corr.	(4) Ownership share	(5) Inputs share
ln Roman eff. distance	-0.0560	-0.212*	-0.0000295	-1.170***	0.399**
	(0.108)	(0.127)	(0.0170)	(0.246)	(0.197)
Joint Roman rule (cent.)	0.707***	0.697***	0.0254**	0.521**	0.939***
	(0.124)	(0.126)	(0.0126)	(0.265)	(0.133)
ln Shipment cost	-2.038***	-2.052***	-0.0828***	-1.912***	0.434***
	(0.0842)	(0.0703)	(0.00880)	(0.149)	(0.0960)
Observations	14,280	14,280	13,806	13,924	14,042

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. The dependent variable are the share of (freight) exports from an origin to a destination NUTS2 region in the total imports of the destination region in values (1) and in weight (2), the night-time light intensity Pearsons correlation (3), the share of firms in destination that are at least $> 25\%$ owned by origin firms (4) and the share of inputs from origin region in the total import of inputs of the destination (5). The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. The specifications also include among the covariates the common country belonging, biome similarity, waterways and Mediterranean dummies as well as the latitude difference. All specifications include destination and origin Fixed effects and use PPML estimates (light corr. has also only positive values). Source: Author's computations on Flückiger et al. (2022), European Commission JRC and Eurostat ERFT data

Table 1.B12: Gravity PPML estimates - Alternative economic ties measurements, without geodesic distance control

	(1) Input Share	(2) Input Share	(3) Input Share	(4) Input Share	(5) Input Share
ln Roman eff. distance	-0.346**	0.562***	0.247	0.129	0.0786
	(0.154)	(0.202)	(0.214)	(0.214)	(0.234)
Joint Roman rule (cent.)		1.035***	1.161***	0.878***	0.903***
		(0.104)	(0.143)	(0.126)	(0.160)
ln Geodesic distance			0.296***	0.606***	0.411**
			(0.0846)	(0.108)	(0.198)
Same State				0.680***	0.750***
				(0.0893)	(0.108)
ln Shipment distance					0.269
					(0.167)
Observations	14042	14042	14042	14042	11191

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. The dependent variable is the share of inputs from an origin to a destination NUTS2 region in the total inputs of the destination region in constant 2013 \$. The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. The specifications also include among the covariates the common country belonging, biome similarity, waterways and Mediterranean dummies as well as the latitude difference. All specifications include destination and origin Fixed effects. Source: Author's computations on Flückiger et al. (2022), European Commission JRC and Eurostat ERFT data

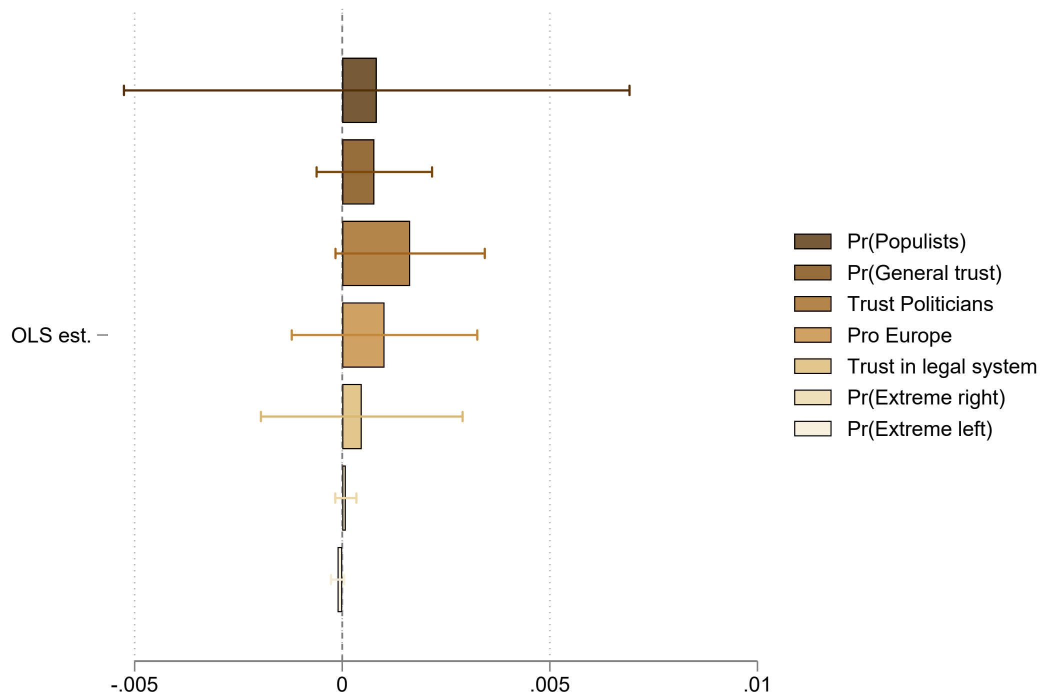
Table 1.B13: Gravity PPML estimates - Roman effective distance and inputs linkages

	(1) Trade Share	(2) Trade Share	(3) Trade Share	(4) Trade Share	(5) Trade Share
Joint Roman rule (cent.)	0.553***	0.629***	0.0618	0.0921	
	(0.125)	(0.134)	(0.129)	(0.131)	
Total joint Past rule					0.0359***
					(0.00844)
ln Roman eff. distance	0.198	0.130		0.348**	
	(0.125)	(0.136)		(0.152)	
GPS abs. difference	-0.115***	-0.0604	-0.0288	-0.0316	-0.0487
	(0.0446)	(0.0419)	(0.0385)	(0.0407)	(0.0411)
ln w. geodesic distance		-0.941***	-0.912***	-0.938***	-0.825***
		(0.155)	(0.157)	(0.156)	(0.143)
ln Geodesic distance	-0.834***				
	(0.129)				
ln Shipment cost	-1.356***	-0.994***	-0.875***	-0.949***	-1.001***
	(0.150)	(0.189)	(0.214)	(0.219)	(0.184)
Same State	1.122***	1.285***	1.404***	1.388***	1.265***
	(0.0895)	(0.0919)	(0.113)	(0.113)	(0.101)
Past states exposures	No	No	Yes	Yes	No
Observations	10,920	10,920	10,920	10,920	10,920

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

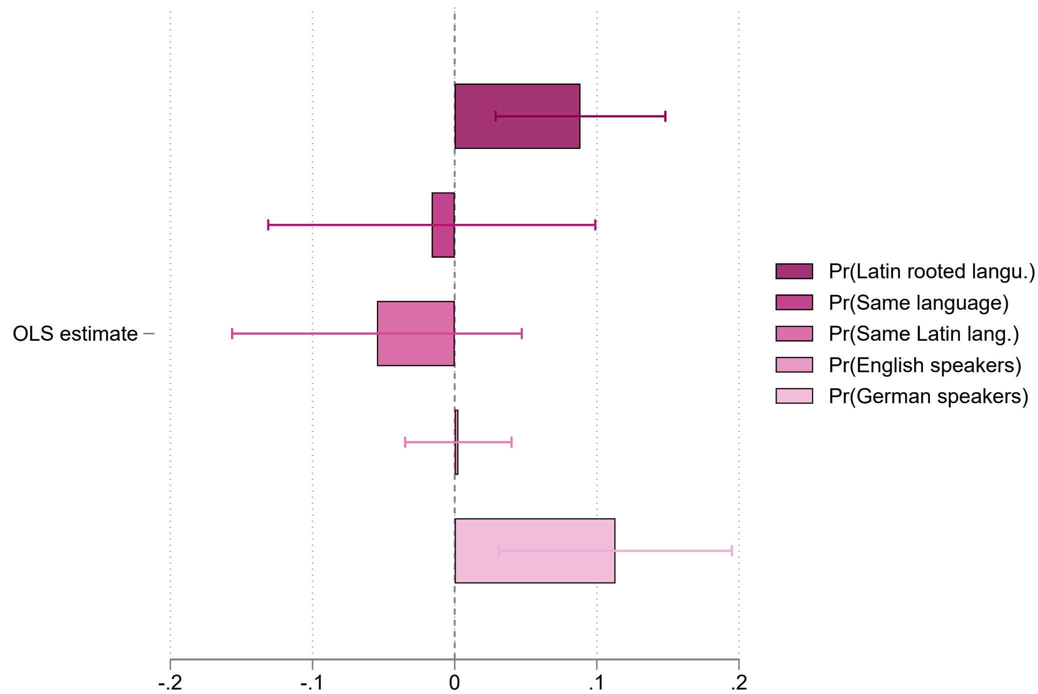
Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. The dependent variable is the share of freight trade from an origin to a destination NUTS2 region in the total freight trade of the destination region in constant 2013 \$. The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. The *GPS abs. difference* is the standardized (divide by sd) regional absolute difference of the first principal component from a PCA realized on the the GPS survey. The specifications also include among the covariates the common country belonging, biome similarity, waterways and Mediterranean dummies as well as the latitude difference. All specifications include destination and origin Fixed effects. Source: Author's computations on Flückiger et al. (2022), *Cliopatra*, European Commission JRC and Eurostat ERFT data

Table 1.B14: PPML estimates - Controlling for Preferences differences (std.)

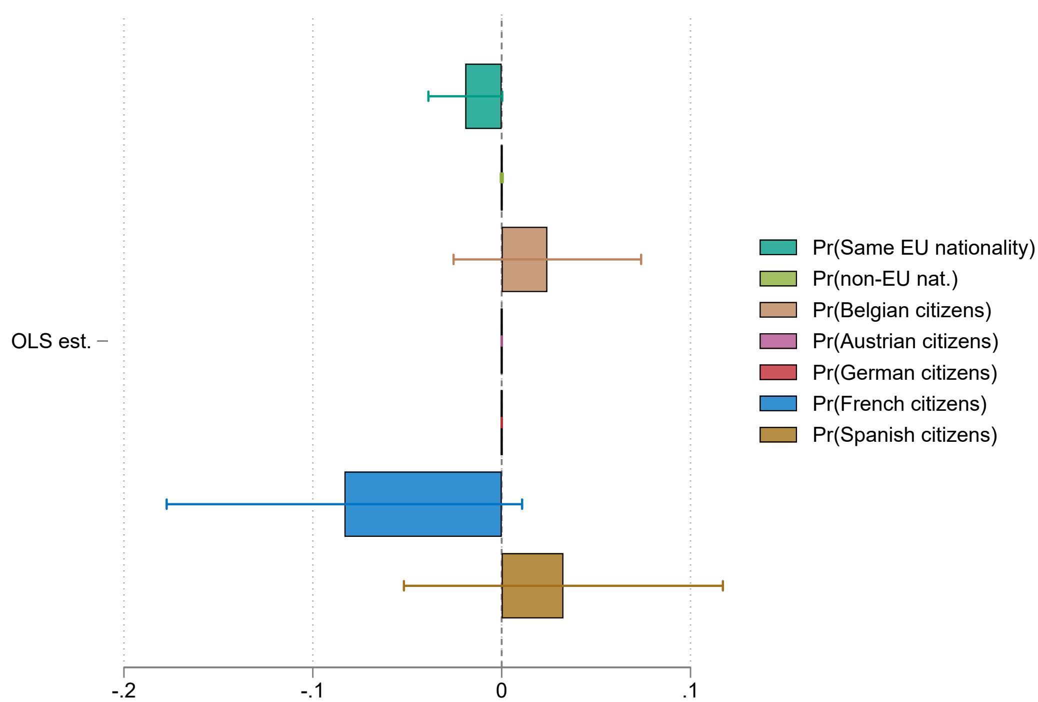


Note: This graph depicts the estimates from the eq. (1.1) specification using tolerance and trust dependent variables. It includes the probability of drawing two people from the respective regions who voted for populist, extreme right or left parties and who claim to trust people in a general way, as well as the interaction of the average trust in the legal system and politicians (sace 0-1). Source: Author's computation on the European Commission datasets and on Algan et al. (2017)'s own computations on the European Social Survey dataset.

Figure 1.B5: OLS estimates - Tolerance, trust, and institutional mechanisms variables without past states exposure controls



(a) (b)



Note: The graphs display the joint centuries under Roman rule estimates and 95% confidence intervals (standard errors clustered at the origin and destination levels) from the eq. (1.1) specifications with (log of) weighted geodesic distances and freight trade cost as geographic connectivity control. The dependent variables are the standardized probabilities of drawing two random persons in each region speaking the considered language (a) or being born in the same defined country. Source: Author's computation on Flückiger et al. (2022), *Cliopatra* and Eurobarometer data.

Figure 1.B6: OLS estimates - Linguistic (a) and migration (b) mechanisms variables with past states exposure controls

1.B.3 On the Roman law, Uncertainty and European law framework

Beyond languages and cultures, similar institutions that promote collaboration and ensure the effective application of legal frameworks are also likely to affect connectedness between communities and decrease uncertainty. Pieces of evidence from Becker et al. (2016) suggest that the Habsburg Empire and its effective bureaucracy created trust in the government by enforcing the rule of law. While the Roman administration was not necessarily as efficient, Roman law is considered one of the main roots of modern civil law in many countries, notably the European territories that were part of the Roman Empire for more than four centuries. Interestingly, Roman law evolved with the changes in Roman society (often emerging from territorial expansions) through continuous interactions between enacted and case law as well as custom, mostly through case law. Specialized jurists compiled and transmitted the content and evolution of this unique system of law until it was codified in later periods (Mousourakis, 2015, pp. VII-VIII). The participation of jurists in the government is already established at the beginning of the Empire and grew in importance with the increasing complexity of issues associated with growing internal trade (Mousourakis, 2015, Chapter 2, pp. 70-83). The 212AC *Constitutio Antoniniana* made Roman law universal for the whole Roman territory and people. These elements imply that regions in the Roman dominion faced less heterogeneity in legal norms than in most of antiquity and the medieval states, and were integrated into a state able to promulgate norms to facilitate trade and adapt to new situations. It is likely to have decreased uncertainty, especially between regions subject to these rules since an early period. Following the German migrations and the official collapse of the Western Empire, the populations faced the coexistence of different norms again; most German states kept the Roman law in application for their Roman population, and sometimes followed its Byzantine modifications, even if they did not have equivalent administrative powers to enforce it (Mousourakis, 2015, Chapter 7, pp. 240-243). As a consequence, the Roman law system may also have affected posterior norms in the long run: German states partially applied Roman law, mostly in areas with a high proportion of Roman populations, and the Church law was also strongly influenced by Roman law. Similar legal frameworks may have decreased uncertainty between regions deeply integrated in the Roman Empire in a persistent way, resulting in stronger trade connections in the long run.

1.B.4 The Roman Empire and the early spread of Christian ideas

In this subsection, I ensure that the main results of the article are not driven by the spread of Christian faith in the Roman Empire and assess the impact of Roman integration on the spread of Christian ideas in this society. As the spread of this global religion is considered to be partially related to past migrations, it allows to test indirectly the past migration and diaspora channels[45] that could have resulted in long-run ties. This subsection is structured as follows: Firstly, I assess the impact of the Roman network and integration, expanding Fousek et al. (2018)'s research using the dependent variables presented in this article. Secondly, I test whether the presence of early Christian networks is associated with past and current trade networks.

The diffusion of Christianity in the Roman Empire

To test if the early diffusion of Christian faith is related to connectivity and political integration under the Roman Empire, I use the progressive apparition of Christian congregations (defined as a community of at least 200 members) in the Roman cities between the first and the fourth centuries from Fousek et al. (2018). It proxies the spread of the new religion in Europe. It allows for testing formally if the Roman political integration and connectivity facilitated the spread of the Christian faith. Focusing on Christian communities existing before the fourth century helps to ensure the results are driven by political integration and networks rather than faith propagation policies[46] following the conversion of the emperor

[45] See Price (2012) for elements suggesting the Christian faith notably spread in the Empire through migrations.

[46] It notably included judicial functions to bishops, direct political influence in the Roman provinces, and public funding.

Constantine (Drake, 2002, Chapter 9). I compute the raw number of centuries these congregations existed before the official collapse of the Western Empire. I then aggregate these values at the regional level by a distance-weighted Epanechnikov[47] average from the exact locations of the congregations. It provides an average early Christianity NUTS2 exposure[48]. While this variable is not easy to interpret in terms of magnitude, it provides a crude measure of early Christian exposure by taking into account both its duration (the period the Christian congregations appeared) and its extent (the number of congregations and their exact locations). A spatial distribution of this variable is available in fig. 1.B7 (a). I use the absolute difference in the Christian exposure between two regions as the main measure of the Christian ideas' spread (measured negatively). To test if connectivity under the Empire shaped its diffusion, I apply the following model:

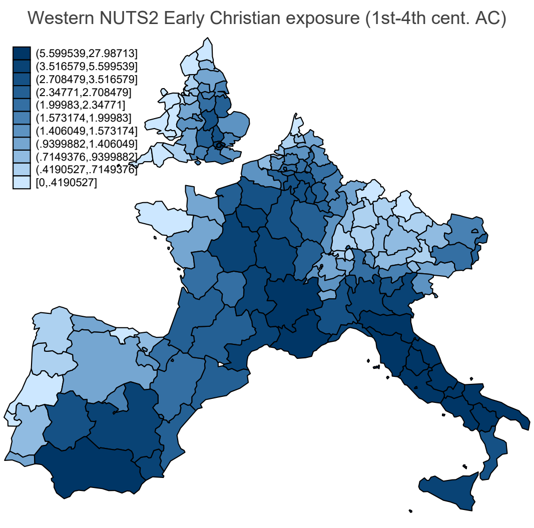
$$|Christ.\ exposure_i - Christ.\ exposure_j| = exp(\omega Roman\ integration_{ij} + \delta ln(Roman\ Distance_{ij}) + \beta ln(Geod.\ Distance_{ij}) + \gamma_i + \gamma_j + \epsilon_{ij}) \tag{1.B2}$$

The PPML estimates are displayed in the table 1.B15 columns (1) and (2). Roman political integration is strongly associated with a lower difference in exposure to Christian ideas, while the Roman effective distance has a negative relationship and becomes insignificant when accounting for latitude and geodesic distances. As the distance radius remains a relatively arbitrary choice without evidence of an accurate Christian congregation's general sphere of influence, I also compute the lags (in centuries) of the apparition of Christian communities among the regions as an alternative extensive measure in columns (3) and (4) as well as the difference in the total number of congregations before 304 AC in columns (5) and (6). All the estimates have the expected signs, but the geodesic distance. It is potentially endogenous

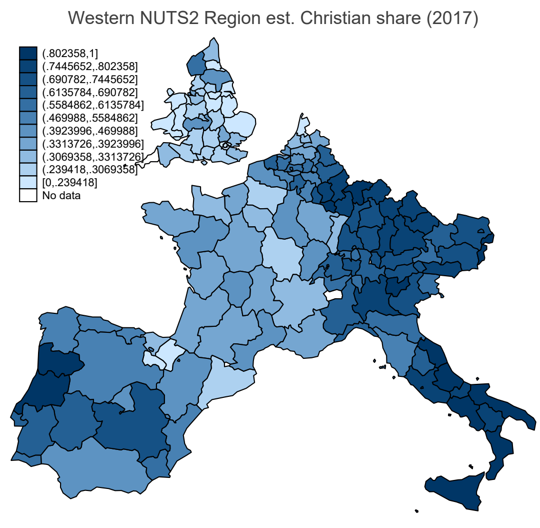
[47]Note: a linear average provides highly similar values and results.

[48]It is technically the average distance-weighted sum for each region of the numbers of centuries each congregation (belonging or being close (100 degrees)) to the region existed before the Western Empire "collapsed". For instance, it would be slightly lower than 4 for a small region with two Christian congregations that appeared in the third century (without congregations in the neighboring regions).

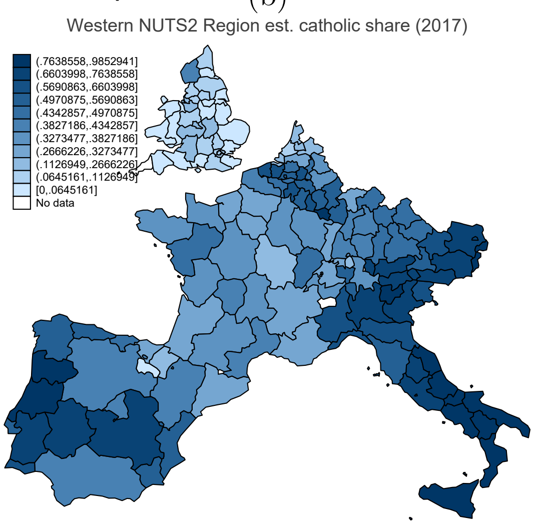
to Roman effective distance in (4) and (6). A decrease in the Roman effective distance is associated with lower time differences in the onset of Christianity, and with a similar intensity of the Christians' presence.



(a)



(b)



(c)

Note: Each map depicts the degree of Christian exposition of the NUTS2 2013 Western European regions, during the Roman Empire era (a) and in 2015-2017 (b). The first (a) is the average distance-weighted sum for each region of the number of centuries the 0-304AC Christian congregations existed before the Empire collapse (fifth century). It was computed as the regional average of a raster created using a 100-degree radius heatmap method. The second (b) shows estimated Christian shares from survey data (2017 European Social Survey, completed by the 2015 Eurobarometer 83.4 for Belgium, Germany, and Luxembourg). Warning: It is important to note that German values were only available at the NUTS1 level and directly applied to NUTS2 level. This decision that produces a biased distribution in favor of a geographic homogeneity for these specific regions was made because of their (small) size and number, which does not compromise the general overview of the map scale. Source: Author's computation on Fousek et al. (2018), ESS, and European Commission datasets and shapefiles.

Figure 1.B7: Regional NUTS2 Christian exposure, 1st-4th century AC and 2017

	(1)	(2)	(3)	(4)	(5)	(6)
	Diff.	Diff.	Diff.	Diff.	Diff.	Diff.
	Christ. exposure	Christ. exposure	Onset Christ.	Onset Christ.	# congregations	# congregations
Joint Roman rule	-0.870***	-0.827***	-0.475***	-0.294**	-0.839***	-0.579***
	(0.0831)	(0.0918)	(0.125)	(0.141)	(0.148)	(0.154)
ln Roman eff. dist.	0.171**	0.0959	0.288**	0.408**	0.616***	0.987***
	(0.0741)	(0.115)	(0.131)	(0.201)	(0.213)	(0.224)
ln Geodesic dist.		0.191**		-0.100*		-0.271***
		(0.0765)		(0.0580)		(0.0783)
Geogr. controls	No	Yes	No	Yes	No	Yes
Observations	20,880	20,880	20,880	20,880	20,880	20,880

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. The dependent variable are the absolute difference between the fig. 1.B7 (a) NUTS2 early Christianity exposure in (1) and (2), the absolute difference in centuries between the respective apparition of the first Christian congregation in (3) and (4), and the absolute difference between the numbers of Christian congregations that existed between the first and the fourth centuries in (5) and (6). The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. The specifications with geographic control specifications include among the covariates the biome similarity, waterways and Mediterranean dummies as well as the latitude difference. All specifications include destination and origin Fixed effects. Source: Author's computations on Flückiger et al. (2022) and Fousek et al. (2018) data.

Table 1.B15: PPML estimates - Roman integration and Christianity diffusion

Early Christianity, cultural convergences and trade

The Christian exposure variable is, however, a relatively poor predictor of current Christians' presence estimated in fig. 1.B7 (b), suggesting that this measurement is more likely to capture past migrations or cultural similarities rather than a contemporary Christian influence. Under this consideration, the cultural convergences detected in fig. 1.4 may have Christian roots. Replications of table 1.5 specification (4) estimates controlling for early Christian diffusion are available in table 1.B15. Including these controls does not strongly affect the magnitude and the significance.

	(1) Trade share	(2) Trade share	(3) Trade share
Christian exposure diff.	0.00703		
	(0.0105)		
Christian onset diff.		0.00401	
		(0.0440)	
# congregations diff.			-0.00749
			(0.0108)
Joint Roman rule	0.387**	0.362**	0.348*
	(0.188)	(0.181)	(0.183)
ln Roman eff. distance	0.231	0.224	0.231
	(0.150)	(0.151)	(0.150)
ln Geodesic distance	-0.658***	-0.647***	-0.651***
	(0.115)	(0.117)	(0.117)
ln Shipment cost	-1.348***	-1.352***	-1.349***
	(0.146)	(0.147)	(0.147)
Geogr. controls	Yes	Yes	Yes
Past states exposure	Yes	Yes	Yes
Observations	13,572	13,572	13,572

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

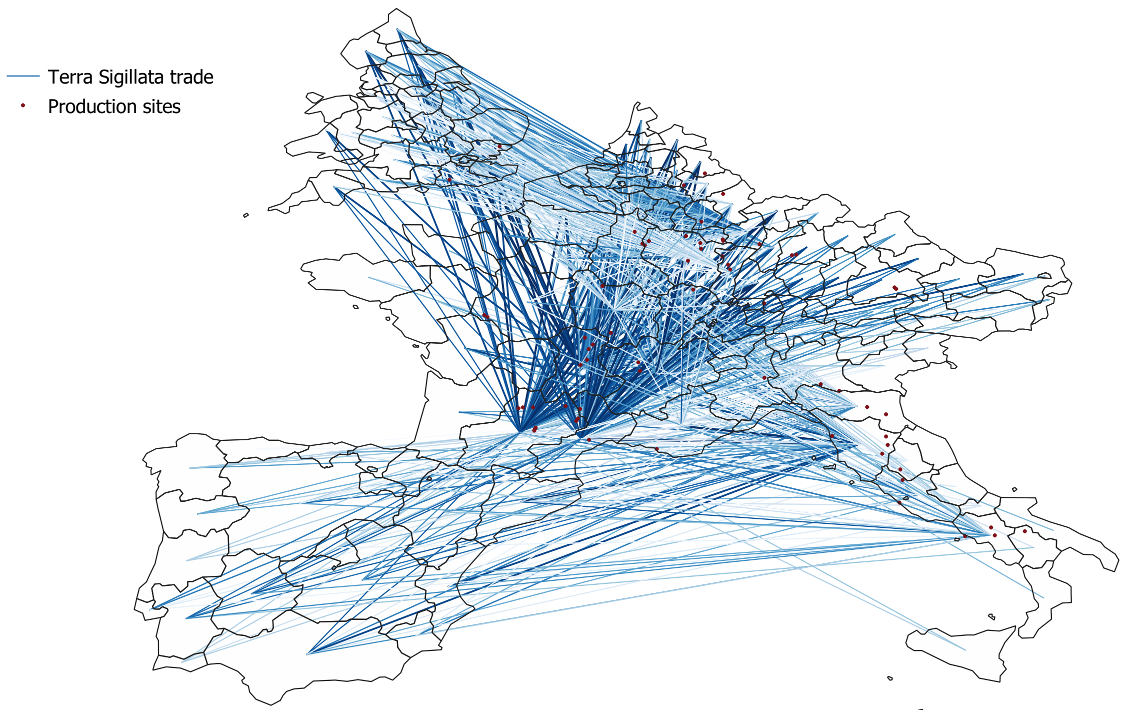
Note: All the models are estimated using Poisson Pseudo Maximum Likelihood (PPML) estimates. Robust se clustered at the origin and destination NUTS2 regions level are reported in parentheses. The main independent variable are the absolute difference between the fig. 1.B7 (a) NUTS2 early Christianity exposure in (1), the absolute difference in centuries between the respective apparition of the first Christian congregation in (2), and the absolute difference between the numbers of Christian congregations that existed between the first and the fourth centuries in (3). The *Joint Roman rule* is the number of centuries the regions jointly belonged to the Roman Empire and Republic. The *Roman effective distance* is the average cost associated with shipping goods along the least cost path between the Regions, given the Roman transport network and Roman-era-specific freight rates for each mode of transport. The specifications with geographic control specifications include, among the covariates the complete past states' exposure variables, the biome similarity, waterways and Mediterranean dummies as well as the latitude difference. All specifications include destination and origin Fixed effects. Source: Author's computations on Flückiger et al. (2022), Joint Research Center, Eurostat, Bennett et al. (2025) Cliopatra and Fousek et al. (2018) data.

Table 1.B16: PPML estimates - Early Christianity diffusion and current trade

1.B.5 The continuous Roman influence in the economic Western linkages

Flückiger et al. (2022) emphasized a continuous influence of past Roman network structure on trade proxies at various periods at the grid level (0.5x0.5 lat. and long. degrees). They capture Roman-era trade through *Terra Sigillata*: The Gallo-Roman *Terra Sigillata* items consist of manufactured red gloss tableware produced massively in centers located in various regions in the West. It covers a wide range of goods and represents approximately a bit more than 10% of the Roman trade at that time (Mees, 2018). Craftsmen working with this material applied stamps to their products. These stamps make it possible to identify the production sites where the objects—whose remains have been found throughout Western Europe—were originally manufactured, allowing Flückiger et al. (2022) to estimate trade flows. The fig. 1.B8 depicts a general overview of these flows' direction and intensity at 2013 NUTS2 level: It draws strong levels of exports from South and central Gaul to Northern regions. These *Terra Sigillata* findings prove that almost, or even potentially, all the European Western regions were involved, at various levels, in *Terra Sigillata* trade with relatively distant producing regions in the Roman Empire.

To assess the continuous influence of the Roman multimodal network on economic ties, Flückiger et al. (2022) use different proxies at various periods of such links: The Black Death onset in the 14th century, the correlation of the evolution of prices between the 13th and the 18th centuries, and financial linkages in 2018. These variables were computed from the authors' direct data collection and external sources indicated in table 1.B17. Estimating gravity equations, they find evidence of a significant influence of the Roman effective distance retrieved from the Roman network (depicted in fig. 1.2) on these variables. In this section, I replicate this part of their work while aggregating these variables at the regional NUTS2 level using the section 1.B.7 methodology. However, I use the early 21st-century trade described in the section 2.4.1 of this article instead of the ownership share. It allows for testing whether the Roman economic and political integrations continuously influenced the



Note: The map depicts Terra Sigillata bilateral trade flows between Western Europe NUTS2 regions that were at least partially under Roman rule. The links are colored according to the size of the flows (the darker, the higher). They are drawn between the geographic centroids of NUTS2 regions. Red dots indicate the centers of production of Terra Sigillata items, signaling exporting regions. Source: Author's computation on Flückiger et al. (2022) and Eurostat shapefiles and data.

Figure 1.B8: Gallo-Roman *Terra Sigillata* bilateral trade flows

intensity of inter-regional trade until the 21st century.

Period	Economic ties proxy	Data sources
Pre Roman-era	Pre-Roman goods findings share	Schauer et al. (2020) Pétrequin et al. (2012)
Roman-era	*Terra Sigillata* items trade intensity (share)	Roman Tableware Dataset
14th century	1346-1351 Black Death onset lags	Christakos et al. (2005)
13th-18th centuries	1208-1790 price correlation	Federico et al. (2021)
21th century	2018 Share of dest. firms owned ($> 25\%$) by origin firms	Orbis

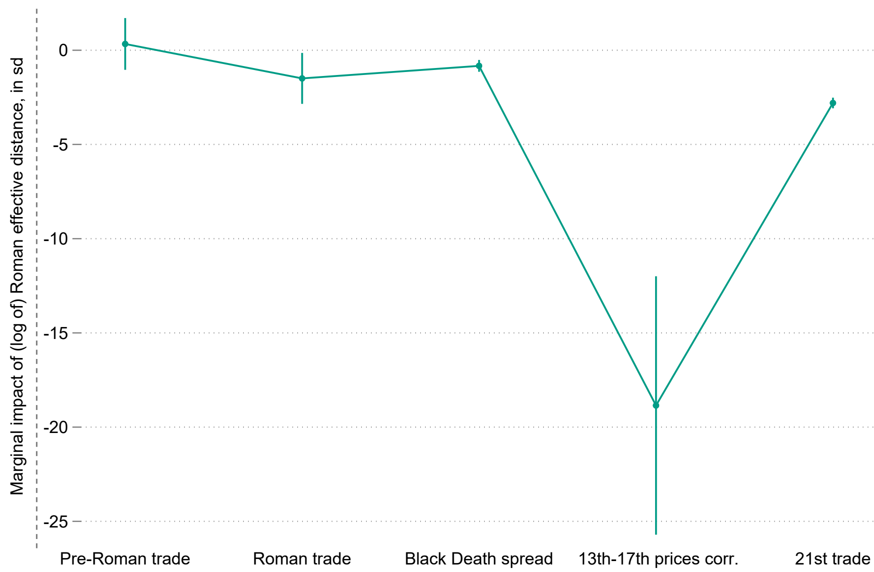
Table 1.B17: Economic linkages proxies used by Flückiger et al. (2022)

As Hillberry and Hummels (2008) showed with US data, estimates related to bilateral economic links may be sensitive to the geographic level of study. Here, I show that the main results are, however, robust to the aggregation at the level of NUTS2 regions. To have a positive way of measuring 14th-century linkages, I define the Black Death spread as follows:

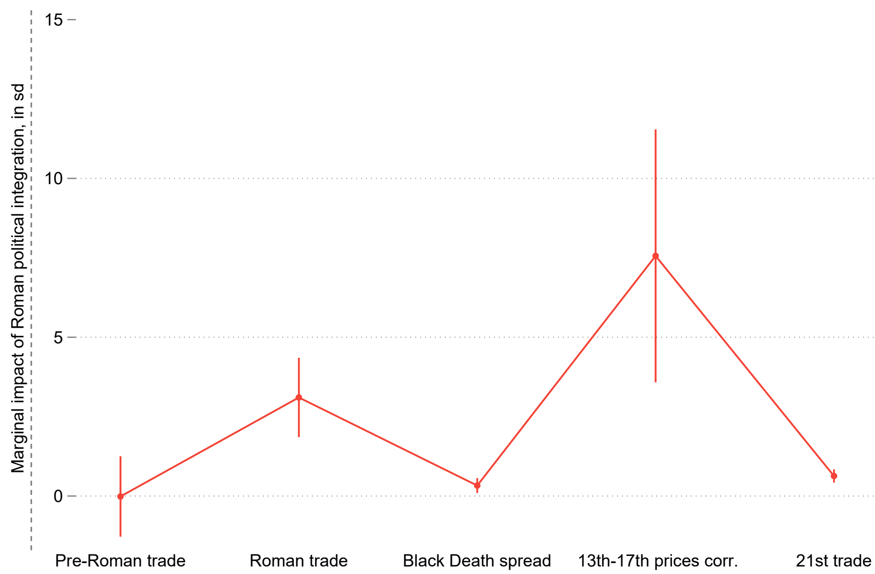
$$Black\ Death\ Spread_{ij} = 26 - (Black\ Death\ lags_{ij}) \tag{1.B3}$$

26 is the maximum number of lags between 2 areas in the Black Death sample. Considering the Black Death expanded partly through commercial ties, this variable proxies how two regions had a stronger market integration at that time than the less connected regions in the sample. It provides a positive measure of the 14th-century connectedness. All the dependent variables are divided by their standard deviation.

The fig. 1.B9 displays the estimates and 95% confidence intervals from the eq. (1.1) gravity equation, excluding 21st-century controls. As the sample coverage varies and the variables are not expressed in the same units or on total trade, the comparison of their magnitudes remains challenging; however, both the Roman effective distance (a) and political integration (b) are associated with a continuous positive and significant influence on the intensity of trade linkages across time. Additionally, the model does not detect any pre-trend associated with these variables (it is worth noting that it may come from a lack of variation due to the restricted sample).



(a)



(b)

Note: The graphs display the (log of) Roman effective distance (a) and the joint centuries under Roman rule (b) estimates and 95% confidence intervals (standard errors clustered at the origin and destination levels) from the eq. (1.1) specifications with (log of) geodesic distances as geographic connectivity control (without controls for the 21st-century network features). The dependent variables x_{ij} used are standardized by dividing them by their standard deviation. They are ordered according to chronological order: 1) Pre-Roman trade consists of goods found that were produced before the Roman era with an identified origin. It covers Alpine jade (Neolithic), British axeheads (Neolithic), and some metal artefacts. 2) Roman trade consists of *Terra Sigillata* goods found that could be associated with a specific Roman production center. 3) Black Death spread is defined in eq. (1.B3) as the difference of lags in the onset of Black Death between two regions and the maximum number of lags in the covered geographic area. 4) 13th-18th price correlations consist of the 1208-1790 regional prices Pearson's correlation. 5) 21st trade consists of the 2011-2019 regional freight trade. All trade variables (1, 2 and 5) are expressed in intensive terms, that is the share of goods coming from an origin in the total goods found in destination. 1, 2, and 5 are associated with PPML estimates, while 3 and 4 could only be estimated through OLS due to the nature of the variables. Source: Author's computation on Flückiger et al. (2022) and Eurostat ERFT data.

Figure 1.B9: Western European economic linkages across time gravity estimates: Roman effective distance (a) and joint centuries under Roman rule (b)

1.B.6 Aggregation sensitivity analysis

This section extends the section 1.5.1 analysis and the work of Flückiger et al. (2022) by exploring the influence of Roman connectivity on trade and financial ties, and by exploring the possibility of an aggregation bias in the main analysis.

Aggregation sensitivity

The section 1.5.1 investigates the role of Roman integration on European economic ties at the NUTS2 pair level, the most disaggregated level where intranational trade can be retrieved. Recent theoretical consideration (Coughlin and Novy, 2021) shows that the level of analysis has an important influence on estimation. This section addresses the potential limitations associated with relying on relatively large spatial units by comparing these estimates to the grid-cell approach employed by Flückiger et al. (2022): financial linkages can indeed be retrieved at both levels. I also aggregate the main variable at the NUTS1 level to observe the general sensitivity of the coefficients to geographic level changes. To do so, I use a baseline that corresponds to eq. (1.1) specification where trade costs variables only include the (log of) geodesic distance, latitude difference, as well as same state, waterways intersection, same biome, and Mediterranean indicators, following Flückiger et al. (2022). To ensure comparison with larger geographic units, I exclude observations located in an aggregated region that is partially outside the Roman Empire. The estimates are pictured fig. 1.B10. This resampling had little influence on their magnitude[49].

Same State variable - Theoretically, the estimated border effect should decrease with aggregation (Coughlin and Novy, 2021), but empirical studies do not perfectly align with this prediction (Havranek and Irsova, 2017). While going from NUTS1 to NUTS2 regions corroborates this element for both trade and financial linkages (fig. 1.B10 (a) and (b)), the grid-cell level shows a stronger coefficient that may be a specificity of financial ties. These

[49]For instance, the elasticity of financial intensity to Roman eff. distance is equal to 0.425 with the 693,838 related grid cells instead of 0.401 with the 731,823 grids full sample.

patterns may also reflect compositional effects: intraregional pairs are outside the sample at aggregated levels. For example, two different grids belonging to the same NUTS2 region mechanically disappear when aggregating. If there is a regional home bias at the NUTS2 level (Santamaria et al., 2023), this disappearance is likely to affect the results. A rationale behind this pattern could be that NUTS2 interregional bilateral ownership with very close neighbors is associated with a lower border effect than distant pairs.

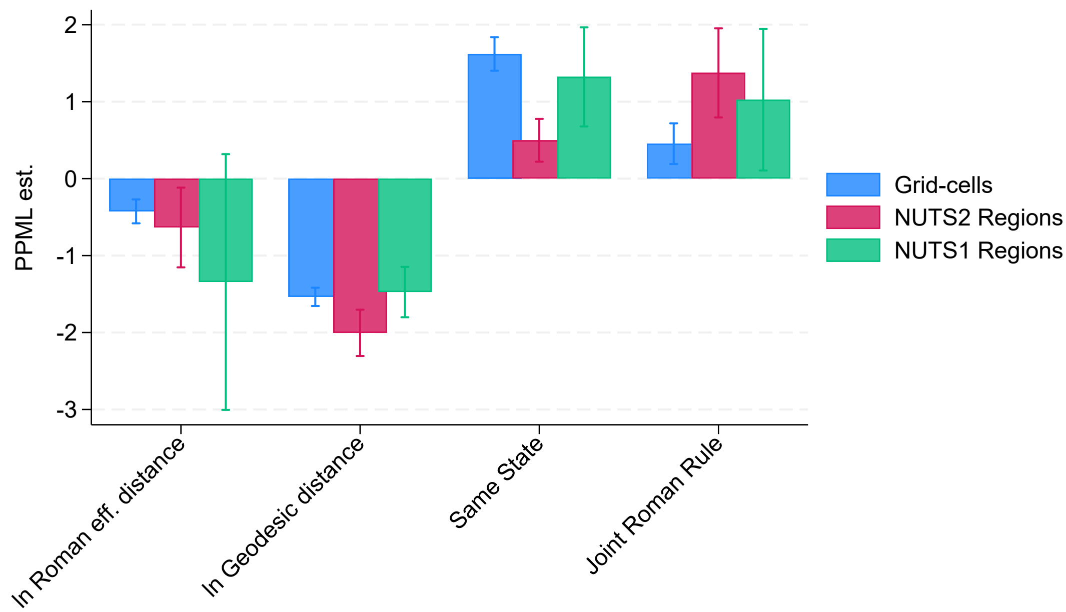
Joint Roman Rule - Interestingly, the Joint Roman rule follows the opposite pattern than the Same State variable with ownership shares: its magnitude is maximized at the NUS2 level and is similar to that of the NUTS1. Both are higher than the grid-cell result, but only NUTS2 is significantly different. Using trade shares, the estimate doubles at the NUTS1 level.

Distance variables - The ownership-share elasticity with respect to geodesic distance in the baseline is 1.5 for grid-cell pairs *versus* 2 for regional NUTS2 pairs. Large geographic units imply larger trade costs (Coughlin and Novy, 2021), and then a stronger sensitivity to trade costs variables. Under this logic, it should rise from NUTS2 to NUTS1 level, with the size of the underlying regions (Hillberry and Hummels, 2008; Coughlin and Novy, 2021). Freight trade estimates align with this prediction. However, estimates suggest otherwise for ownership ties, which may be rooted in the difference between financial and trade ties.

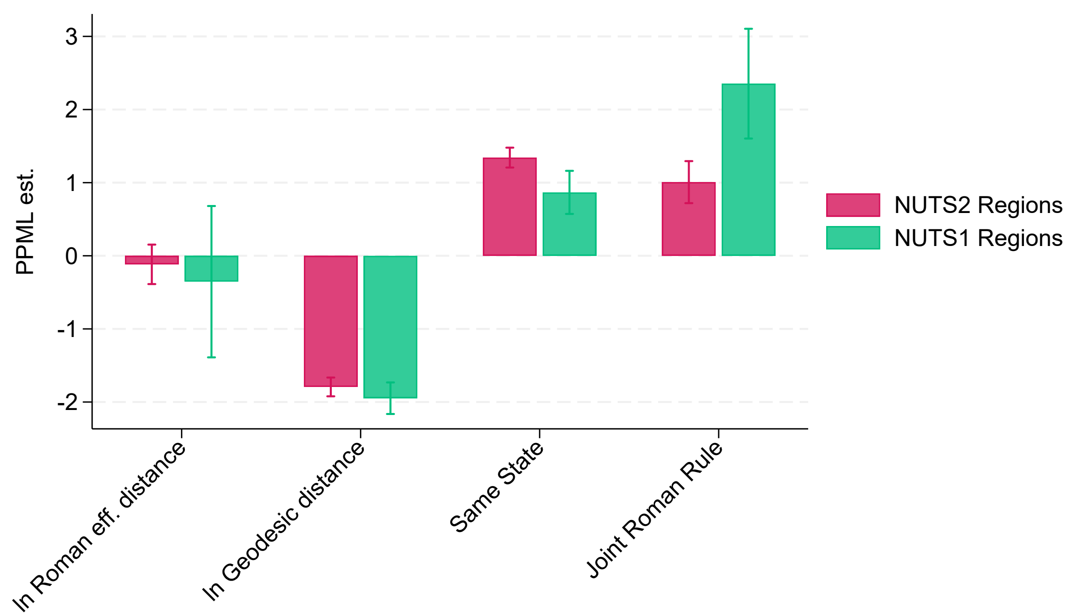
Together, these results reveal a relatively strong and non-monotonic sensitivity of the estimated (semi-)elasticities to aggregation. Compositional differences, as aggregation means the loss of intraregional information, as well as the differences between financial and trade ties, are a plausible explanation behind the non-linear patterns observed.

Roman connectivity and long-run economic linkages

The section 1.5.1 investigates the role of Roman historical political integration on European economic ties beyond geographic connectivity and Roman infrastructure. This research



(a)



(b)

Note: The plot pictures PPML estimates and 95% Confidence intervals of a gravity equation with (a) ownership shares (the shares of firms in a destination belonging to an origin firm (> 25%)) or (b) trade shares as the dependent variable. It replicates Table 3, second column specification of Flückiger et al. (2022) that includes origin and destination fixed effects, latitude difference, as well as Mediterranean, biome similarity, and Waterways dummies among the covariates. The gravity equation is estimated at the grid-cell level defined by Flückiger et al. (2022), NUTS2, and NUTS1 levels using observations common to the present research (It excludes grids belonging to a region partially located in the Roman Empire's maximum extension).

Figure 1.B10: PPML Gravity estimates - Financial (a) and trade (b) linkages at the grid-cell, NUTS2 and NUTS1 levels of analysis

also provides an opportunity to reassess the long-run influence of the Roman multimodal transport network on contemporary economic connections using regional freight trade data; More particularly, by comparing trade and financial linkages: Pearson's correlation between the two measures is equal to 0.57, which is weaker than expected, and their relationship is slightly monotonic (section 1.B.6 fig. 1.B2). Additionally, the aggregation sensitivity analysis performed in the previous section presented distinct patterns for both measures. This section replicates the analysis of Flückiger et al. (2022) to investigate the robustness of their findings to trade ties in table 1.2.

When investigating the influence of early Roman connectivity on 2011-2019 interregional NUTS2 trade linkages, the table 1.2 second column estimates a (conditional) elasticity of trade intensity to Roman effective distance equal to 0.617. It replicates the specification of Flückiger et al. (2022), who obtained an elasticity of 2018 financial linkages intensity of 0.401, working at a 0.5x0.5 degree grid-cell pair level. This difference is related to the levels of analysis and not the different nature of these variables: Aggregating their ownership linkages at the NUTS2 pair level results in an elasticity of 0.612 (section 1.B.6 fig. 1.B10), very close to the estimated trade elasticity[50].

Interestingly, the coefficient becomes insignificant once controlling for the current European transport network, either through the average revealed bilateral freight trade distance in table 1.2 col. (4) or through estimated bilateral freight trade costs in column (6). While this second result is driven by the exclusion of Switzerland from the sample[51], the comparison between table 1.2 col. (3) and (4) show that insignificance is achieved when controlling for actual truck distance. These results show that the Roman network influence on western interregional trade is not robust to resampling and decreases greatly when accounting for the current European road network. Surprisingly, however, regressions on ownership share present significant estimates in table 1.3 col. 5 despite the same resampling and the inclusion

[50] Note that the influence of the Roman transport infrastructure on historical economic connections is also still persistent across time at the NUTS2 level (section 1.B.5 fig. 1.B9 (a)).

[51] The section 1.B.2 table 1.B4 Replicating table 1.2 col. (2) while excluding Switzerland and Liechtenstein shows insignificant results.

of such controls[52], consistent with the conclusions of Flückiger et al. (2022).

Intuitively, the source of the divergence may arise from the fact that transport costs are less likely to influence financial ties; they indeed do not necessarily require the transportation of goods if investments are horizontal and are associated with different kinds of costs. The changes in the estimates of trade variables in table 1.2 highlight their sensitivity to transport costs, while the ownership share has an insignificant positive estimate[53]. Trade and financial linkages may therefore differ slightly as pictured in fig. 1.B2. As a consequence, the regressions are unable to detect a Roman network influence on current trade linkages other than through the current network, following the ancient Roman one.
Flückiger et al. (2022) computed distances from a Gabriel graph[54] drawn from the main Roman battles. Using it to instrument the Roman effective distance, they intended to capture mostly variations arising from Roman military purposes. The results remain robust to this strategy as attested in section 1.B.2 table 1.B8.

The dichotomy of the results according to trade or financial linkages raises several concerns. It does not seem to arise from a geographical aggregation bias, as it seems to have the opposite pattern for Roman effective distance (fig. 1.B10). Moreover, the network in Switzerland, and especially its connections to other countries, is partly determined by its mountains and geography, which may make geodesic distance an imperfect proxy of trade costs with its regions. Taken together, these results suggest that the Roman connectivity influence on trade, if it exists, is mostly driven by the current network following the old Roman network. On the other hand, its robust influence on financial linkage patterns suggests that trade and financial costs are fundamentally different objects.

[52] A more complete replication following the table 1.2 specifications is available in section 1.B.2 table 1.B7, achieving the same conclusion.

[53] When controlling for geodesic distance; otherwise, the elasticity to transport costs is approximately 1.91 in section 1.B.2 table 1.B12, which is close to the trade share elasticity.

[54] A Gabriel graph is a geographic network built from a set of points (here Roman battles) based on the idea that points are linked if there is no third point closer to both of them. If a third point falls inside the diameter circle, then the direct connection between the two points is considered redundant and is removed (Matula and Sokal, 1980)

1.B.7 Weight appendix

The variables from Flückiger et al. (2022) are aggregated at the NUTS2 regional level using an area-weighted spatial average formula adapted to their definition. The same methodology is applied for political variables from the *Cliopatra* data. The weights are based on the geographic areas of the grid cells and the regions. For instance, aggregated distances between two regions would be the average distance between two cells belonging to these regions. The aggregated number of firms in a region would be the sum of the firms in its composing cells, assuming a spatially uniform distribution of firms within the cells. These aggregation weights do not proxy for the endogenous theoretical weights to retrieve exact aggregated trade costs and economic flows (Coughlin and Novy, 2021). They are unobservable and assumed equal here. Instead, these spatial weights account for the fact that grid and NUTS2 regions are not a perfect match and that a cell partially lying in a NUTS2 region should have a lower contribution than a cell fully integrated into the region. For economic variables, it relies on the assumption of a uniform spatial distribution of economic activities within cells. This method typically performs well for geographic variables that mechanically respect this uniform distribution hypothesis: The Pearson correlation between the geodesic distance using this method from cells' centroids and its direct computation from NUTS2 centroids is 0.999.

Unilateral variables

Consider grid cells, indexed k, and wider NUTS2 regions, indexed i. First, I want to define a weight to aggregate a variable defined at the grid-level (the number of firms, for instance) X_k at its regional level i. Let $\mathcal{R}$ be the sample set of NUTS2 regions used for the analysis, indexed by i, j. Grid cells and regions are not perfectly nested: a cell k may overlap several regions. Consider the intersection area between cell k and region i:

$$a_{ik} = \text{area}(k \cap i) \tag{1.B4}$$

The sum of these areas in a region i gives the regional area for any regions entirely covered by the sample of cells, and that no region is smaller than a grid cell[55]: $\text{area}_i = \sum_{k \in i} a_{ik}$. Assume the variable X (economic activities in the case of the number of firms) is spatially distributed uniformly within each grid cell. This assumption is less costly whenever cells are small relative to regions, so that the share of cells belonging to two or more regions remains negligible. Then, I define two geographical weights; first, the geograhic share of cell k belonging to region i, that will be used to sum variables:

$$w_{ik}^{\text{sum}} = \frac{a_{ik}}{\text{area}(k)} \tag{1.B5}$$

It is mechanically equal to 1 if the grid lies entirely in the region. Second, the average weight, used when computing regional averages, is:

$$w_{ik}^{\text{sum}} = \frac{a_{ik}}{\text{area}(k)} \tag{1.B6}$$

$$w_{ik}^{\text{avg}} = \frac{a_{ik}}{\text{area}(i)} \tag{1.B7}$$

As pictured in fig. 1.B11, the value of this first weight is 0.6827 for the 706th cell part that belongs to the *Champagne-Ardenne* NUTS2 region. The weights w_{ik}^{avg} for the *Cliopatra* political integration variables are, in an equivalent way, the share of the intersection between the NUTS2 regions and the past states in the total area of the region.

These two families of weights satisfy natural normalization properties: For the sum weight, the shares of a given cell across all regions it overlaps sum to at most one: $\sum_{i\,:\,k \cap i} w_{ik}^{\text{sum}} = 1$. For the average weight, the cell contributions to a given region sum to one for any region wider than a cell: $\sum_{k\,:\,k \cap i} w_{ik}^{\text{avg}} = 1$. The two weights differ in their normalization reference: w_{ik}^{sum} is normalized by the area of the cell, and is therefore appropriate

[55]There are two exceptions in the sample.

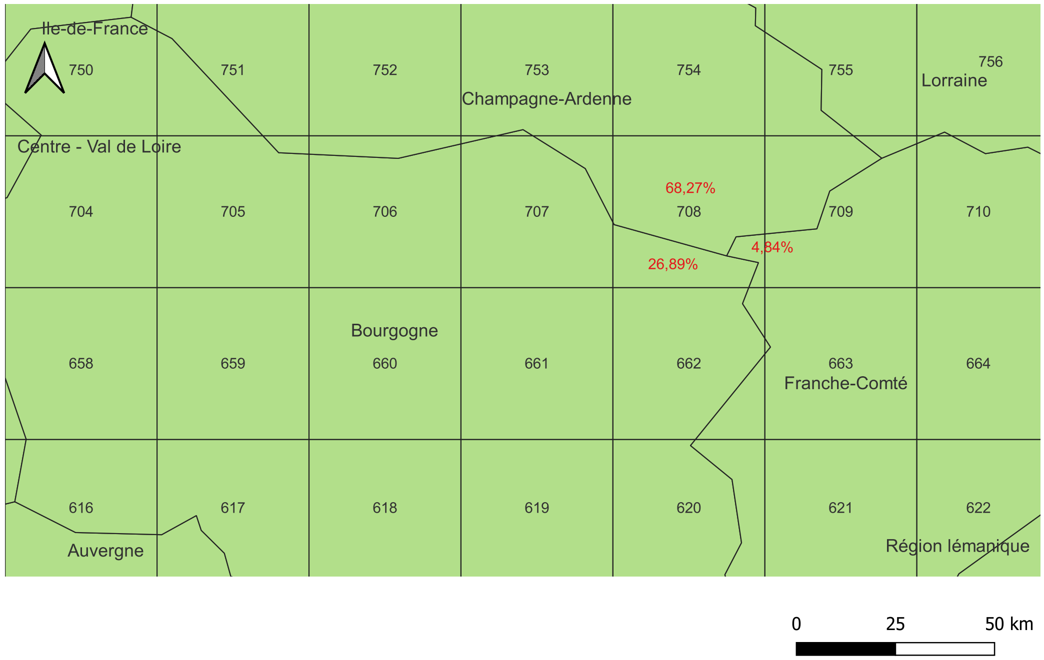


Figure 1.B11: Contribution of the 708th cell to the French regions Champagne-Ardenne, Bourgogne, and Franche-Comté

when the quantity of interest is a stock to be allocated across regions (e.g., number of firms). By contrast, w_{ik}^{avg} is normalized by the area of the region, and is appropriate when computing a regional average of an intensive variable (average elevation, for instance). The following weighted sum can then be applied with the relevant weighting to retrieve the aggregated variable of interest:

$$X_i = \sum_{k \in i} w_{ik} X_k \tag{1.B8}$$

This equation can be used to compute two variables: 1- The (weighted by area) **regional average value of the variable X among the grid cells lying in the region r**. It happens if w_{ik} is defined by eq. (1.B7) (the share of the part of the grid k that belongs to the region i in the total area of the region). Consider the 706th cell part that belongs to the *Champagne-Ardenne* NUTS2 region in fig. 1.B11, and X as the number of firms: If the 706th cell had 100 firms, 68.27 of its firms would be attributed to *Champagne-Ardenne*, 26.89 to *Bourgogne*, and the rest to *Franche-Comté* through the eq. (1.B8) formula. 2- the (weighted by area) **sum of the variable X at the regional level**. It happens if w_{ik} is defined by eq. (1.B6): the ratio of the intersection of the grid cell k with the region i and the grid cell k areas.

Bilateral variables

Consider o (d) an origin (destination) grid cell belonging partially or totally to a wider area i (j). The previous weighted average or sum can be extended to bilateral objects. the aggregation to a pair of regions (i, j) takes the following form:

$$X_{ij} = \sum_{o \in i, d \in j} w_{io} w_{jd} X_{od} \tag{1.B9}$$

As in the unilateral case, two choices of weights yield two distinct aggregated quantities:

Bilateral regional average - Setting $w_{io} = w_{io}^{\text{avg}}$ and $w_{jd} = w_{jd}^{\text{avg}}$ in (1.B9) yields the area-weighted average value of X_{od} across all pairs of cells belonging to regions i and j respectively. This is the appropriate aggregation for distance variables, as the average weighted distance between cells is representative of its region for bilateral variables, such as geodesic distance. To illustrate this one point, let's consider two points: $\mathbf{p} \in i$ and $\mathbf{q} \in j$ are two randomly selected points from two different regions i and j, therefore distributed uniformly over their respective regions. We want to retrieve the expected value of a variable specific to these two points (geodesic distance, for instance) $X(\mathbf{p}, \mathbf{q})$, that is representative for the two regions:

$$\mathbb{E}\left[X(\mathbf{p}, \mathbf{q})\right] = \frac{1}{\text{area}(i)\,\text{area}(j)} \iint_{i \times j} X(\mathbf{p}, \mathbf{q})\, d\mathbf{p}\, d\mathbf{q} \tag{1.B10}$$

Discretizing the domain via the grid cells, each cell o overlapping region i contributes a mass proportional to a_{oi}. The previous expression can then be approximated as:

$$\mathbb{E}\left[X(\mathbf{p},\mathbf{q})\right] \approx \sum_{\substack{o\,:\,o\cap i\\ d\,:\,d\cap j}} \frac{a_{io}}{\text{area}(i)} \cdot \frac{a_{jd}}{\text{area}(j)} \cdot X_{od} \tag{1.B11}$$

which is (1.B9) with the average weights. The lower the grids with respect to the region, the more this equality holds. Under a uniformly spatial distribution, this bilateral regional average is an unbiased estimator of the expected value of X_{od} for two points randomly drawn from regions i and j, respectively.

Bilateral regional sum - Setting $w_{io} = w_{io}^{\text{sum}}$ and $w_{jd} = w_{jd}^{\text{sum}}$, equation (1.B9) yields the area-weighted sum of a variable X_{od} at the regional level. This is the appropriate aggregation for extensive bilateral variables, such as bilateral terra sigillata trade flows.

Bilateral variables with a unilateral normalization reference

A variable X_{od} may be bilateral but normalized according to a reference, either the origin or the destination. For instance, if N_{od} is the total number of firms in the destination grid cell that origin grid cell firms own, then we can express the share of these firms with respect to the number of firms in the destination grid cell N_d:

$$ownership_{od} = \frac{N_{od}}{N_d} \tag{1.B12}$$

Using the eq. (1.B9) aggregation method would then be inappropriate as the reference is destination-specific. Instead, one can directly apply this sum aggregation to the variable N_{od} and apply the destination-specific aggregation to N_d. Applied to the financial linkages, this method approximates the share of the firms in the destination region ($> 25\%$) owned by the origin region firms. Again, an exact equality relies on a uniform spatial distribution of the firms within the grid cells located on the border between different regions. Due to the small size of the grid cells with respect to NUTS2 regions, violations of this hypothesis should have a limited impact.

1.B.8 Sample appendix

Figure 1.B12: Western Roman Empire sample and 2013 NUTS2 European Regions

Industry id.	Label	trade sample
1	Products of agriculture, hunting, and forestry; fish and other fishing products	Yes
2	Coals and lignite, crude petroleum and natural gas	No
3	Metal ores and other mining or quarrying products, peats, uranium and thorium ores	Yes
4	Food products, beverages, and tobacco	Yes
5	Textiles and textile products, leather and leather products	Yes
6	Wood and products of wood and cork (except furniture), pulp, paper, printed matter, and recorded media	Yes
7	Coke and refined petroleum products	Yes
8	Chemicals, chemical products, and man-made fibers, rubber and plastic products, nuclear fuel	Yes
9	Other non-metallic mineral products	Yes
10	Basic metals, fabricated metal products, except machinery and equipment	Yes
11	Machinery and equipment n.e.c., communication equipment, medical, precision and optical instruments, watches and clocks	Yes
12	Transport equipment	Yes
13	Furniture, other manufactured goods n.e.c.	Yes
14	Secondary raw materials, municipal wastes and other wastes	No
15	Mail, parcels	No
16	Equipment and material utilized in the transport of goods	No
17	Goods moved in the course of household and office removals	No
18	Grouped goods	No
19	Unidentifiable goods: goods which for any reason cannot be identified and therefore cannot be assigned to groups 01-16	No
20	Other goods n.e.c.	No

Table 1.B18: NST 2007 Industries in the ERFT survey

Table 1.B19: NUTS2 Sample list and Roman political exposure

NUTS2 Id	**Region name**	**Av. # years under Roman rule**	**Share of Eur. freight imports from ex-imperial regions**
AT11	Burgenland	386	.8507977
AT21	Kärnten	386.0381	.7748395
AT22	Steiermark	386	.8650042
AT32	Salzburg	386	.9092109
AT33	Tirol	408.083	.9488898
AT34	Vorarlberg	397.9127	.9637914
BE10	Brussels Hoofdstedelijk Gewest	450	.9041027
BE21	Prov. Antwerpen	450	.9315559
BE22	Prov. Limburg (BE)	450	.9106925
BE23	Prov. Oost-Vlaanderen	450	.9205996
BE24	Prov. Vlaams-Brabant	450	.9411396
BE25	Prov. West-Vlaanderen	450	.9368142
BE31	Prov. Brabant Wallon	450	.9599512
BE32	Prov. Hainaut	450	.9593045
BE33	Prov. Liège	450	.9249076
BE34	Prov. Luxembourg (BE)	450	.9563347
BE35	Prov. Namur	450	.9563897

The table continues at the next page

Table 1.B19 – continued

NUTS2 Id	Region name	Av. # years under Roman rule	Share of Eur. freight imports from ex-imperial regions
CH01	Région lémanique	452.3304	.9989029
CH02	Espace Mittelland	430.5406	.9973686
CH03	Nordwestschweiz	432.8149	.9955819
CH04	Zürich	432.7565	.9974378
CH05	Ostschweiz	443.1671	.9964136
CH06	Zentralschweiz	415.3368	.9974366
CH07	Ticino	590.3685	.9994487
DE12	Karlsruhe	374.9903	.8758266
DE13	Freiburg	388.7945	.9127239
DE14	Tübingen	386.3922	.8940181
DE21	Oberbayern	378.1958	.8730859
DE27	Schwaben	386	.8728123
DEB2	Trier	450	.8764246
DEB3	Rheinhessen-Pfalz	448.8114	.8670673
DEC0	Saarland	450	.8915813
ES11	Galicia	475	.9926851
ES12	Principado de Asturias	440.1548	.996368
ES13	Cantabria	428.8316	.994572
ES21	País Vasco	470.1836	.9820126

The table continues at the next page

Table 1.B19 – continued

NUTS2 Id	**Region name**	**Av. # years under Roman rule**	**Share of Eur. freight imports from ex-imperial regions**
ES22	Comunidad Foral de Navarra	480.8241	.9815928
ES23	La Rioja	476.1151	.9838384
ES24	Aragón	531.1849	.982786
ES30	Comunidad de Madrid	527.0688	.9744714
ES41	Castilla y León	486.6622	.9901822
ES42	Castilla-La Mancha	536.0872	.9826962
ES43	Extremadura	553.4705	.9931824
ES51	Cataluña	553.1577	.9233134
ES52	Comunidad Valenciana	577.7814	.9885091
ES61	Andalucía	602.0508	.9959473
ES62	Región de Murcia	600.6995	.9897578
FR10	Ile-de-France	450	.9649839
FRB0	Centre - Val de Loire	450	.980854
FRC1	Bourgogne	450	.9744115
FRC2	Franche-Comté	450.3295	.9745625
FRD1	Basse-Normandie	450	.9865357
FRD2	Haute-Normandie	450	.986173
FRE1	Nord-Pas de Calais	450	.9685513

The table continues at the next page

Table 1.B19 – continued

NUTS2 Id	Region name	Av. # years under Roman rule	Share of Eur. freight imports from ex-imperial regions
FRE2	Picardie	450	.9785479
FRF1	Alsace	441.1133	.9551758
FRF2	Champagne-Ardenne	450	.9812029
FRF3	Lorraine	450	.9614877
FRG0	Pays de la Loire	450	.9896005
FRH0	Bretagne	450	.990233
FRI1	Aquitaine	451.017	.9932504
FRI2	Limousin	450	.9894968
FRI3	Poitou-Charentes	450	.991643
FRJ1	Languedoc-Roussillon	518.2689	.9917659
FRJ2	Midi-Pyrénées	487.5585	.9925354
FRK1	Auvergne	450	.9844457
FRK2	Rhône-Alpes	479.5409	.9782084
FRL0	Provence-Alpes-Côte d'Azur	502.8997	.9903243
ITC1	Piemonte	564.9875	.9526315
ITC2	Valle d'Aosta/Vallée d'Aoste	495.3185	.9720844
ITC3	Liguria	618.1832	.9737467

The table continues at the next page

Table 1.B19 – continued

NUTS2 Id	**Region name**	**Av. # years under Roman rule**	**Share of Eur. freight imports from ex-imperial regions**
ITC4	Lombardia	618	.9392956
ITF1	Abruzzo	672	.9374805
ITF2	Molise	672	.9609629
ITF3	Campania	672	.9579408
ITF4	Puglia	672	.9659114
ITF5	Basilicata	672	.9586706
ITF6	Calabria	672	.9779399
ITG1	Sicilia	641	.9922049
ITH1	Provincia Autonoma di Bolzano/Bozen	520.4318	.9356113
ITH2	Provincia Autonoma di Trento	617.8603	.9491443
ITH3	Veneto	605.8564	.8818806
ITH4	Friuli-Venezia Giulia	594.5285	.6135333
ITH5	Emilia-Romagna	624.017	.9500789
ITI1	Toscana	667.3139	.9646295
ITI2	Umbria	672	.9595081
ITI3	Marche	672	.9411059
ITI4	Lazio	672	.9380593

The table continues at the next page

Table 1.B19 – continued

NUTS2 Id	**Region name**	**Av. # years under Roman rule**	**Share of Eur. freight imports from ex-imperial regions**
LI00	Liechtenstein	386	.9736352
LU00	Luxembourg	450	.9943635
NL31	Utrecht	448.9631	.8478159
NL33	Zuid-Holland	450	.8963338
NL34	Zeeland	450	.9081953
NL41	Noord-Brabant	450	.8924124
NL42	Limburg (NL)	450	.9031501
PT11	Norte	480.1263	.9924806
PT15	Algarve	537.1858	.9963092
PT16	Centro (PT)	534.8959	.9952756
PT17	Área Metropolitana de Lisboa	535	.9837375
PT18	Alentejo	540.4597	.9957597
UKC1	Tees Valley and Durham	357	.9739739
UKD3	Greater Manchester	357	.9733844
UKD4	Lancashire	357	.9642494
UKD6	Cheshire	357	.991687
UKD7	Merseyside	357	.975574

The table continues at the next page

Table 1.B19 – continued

NUTS2 Id	**Region name**	**Av. # years under Roman rule**	**Share of Eur. freight imports from ex-imperial regions**
UKE1	East Yorkshire and Northern Lincolnshire	357	.98435
UKE2	North Yorkshire	357	.9821942
UKE3	South Yorkshire	357	.9907014
UKE4	West Yorkshire	357	.9907808
UKF1	Derbyshire and Nottinghamshire	357	.9946988
UKF2	Leicestershire, Rutland and Northamptonshire	357	.9930152
UKF3	Lincolnshire	357	.9956935
UKG1	Herefordshire, Worcestershire and Warwickshire	357	.9849916
UKG2	Shropshire and Staffordshire	357	.9808557
UKG3	West Midlands	357	.9894667
UKH1	East Anglia	357	.9927065
UKH2	Bedfordshire and Hertfordshire	357	.9980263
UKH3	Essex	357	.9942949

The table continues at the next page

Table 1.B19 – continued

NUTS2 Id	Region name	Av. # years under Roman rule	Share of Eur. freight imports from ex-imperial regions
UKI1	Inner London	357	.9969974
UKI2	Outer London	357	.9993193
UKJ1	Berkshire, Buckinghamshire and Oxfordshire	357	.9962036
UKJ2	Surrey, East and West Sussex	357	.9985194
UKJ3	Hampshire and Isle of Wight	357	.9933804
UKJ4	Kent	357	.9931778
UKK1	Gloucestershire, Wiltshire and Bristol/Bath area	357	.9965715
UKK2	Dorset and Somerset	357	.9964966
UKK3	Cornwall and Isles of Scilly	357	.9998704
UKK4	Devon	357	.9966156
UKL1	West Wales and The Valleys	357	.9952232
UKL2	East Wales	357	.9847732

Chapter 2

Birthplace diversity and wages: Evidence from new US migration flows

The increasing arrival of people born in various countries towards high-income destinations raises the issue of the coexistence and the exchange of different norms and ideas that affect workers' productivity. A review of the literature reveals essentially weakly positive or null effects of regional diversity on average wages. Using US data between 2000 and 2019, this article goes to more local levels to study the impact of diversity through interactions and local labor market mechanisms. It shows that metropolitan immigrant diversity of low-skilled workers from the same sectors and occupations is positively related to their wages, while high-skilled workers are unaffected. It also provides evidence of a local positive effect of migrant diversity among low-skilled workers with similar experience and education that captures these complementarities. Relying on a shift-share instrumental strategy, this approach exploits the skill similarities and the high levels of interaction of workers from those groups to disentangle the diversity effects from general skill heterogeneity related to national specialisations. Finally, it exploits the rich compositions of the early 21st century's migration flows to consider the influence of cultural and economic heterogeneity in origins. The gains are maximized when immigrant birthplace diversity in a local occupation is the product of inflows from countries close economically and culturally to the US population, and from culturally distant countries for diversity at the level of workers similar in education and experience. The results, mostly driven by workers without a college education, suggest that communication costs may cancel complementary benefits from diversity locally, but not when occupational mobility is taken into account.

2.1 Introduction

The increasing arrival of people from different origins in high-income countries raises the issue of the coexistence and the exchange of different norms and ideas that affect local economic outcomes. Different disciplines consider different mechanisms behind such effects; most of them are built on the idea that the birthplace of human beings has a lasting impact on them through differences in education, local norms, and culture. Consequently, the geographic gathering of people associated with different birthplaces may result in the exchange of ideas (Hong and Page, 2001), the complementarity between people from different backgrounds, and greater freedom of innovation by weakening dominant norms (Kets and Sandroni, 2021), and also in a lack of unity, conflicts, and communication issues (Richard et al., 2002; Dale-Olsen and Finseraas, 2020). These are likely to influence economic performance and innovation (Ozgen, 2021; Docquier et al., 2020; Ozgen et al., 2017; Parrotta et al., 2014).

The study of this phenomenon has been the subject of a relatively substantial number of works that belong to a long tradition sometimes referred to as the *economics of diversity* (Stirling, 1998; Polachek et al., 2006; D'Ippoliti, 2011; Nijkamp et al., 2015) that assesses the economic consequences of compositional diversity among Human communities[1], notably through migration patterns (Alesina et al., 2016). A considerable strand of this literature notably studies the influence of immigrant diversity on local labor markets or economic outputs, considering diversity at local geographic levels such as administrative regions or cities, with relatively mixed results (Kemeny, 2017; Ozgen, 2021). This approach is likely to give a strong weight to broad complementarities between skills embodied by specific origins and less to local interactions among workers. On the other hand, other contributions study the influence of diversity directly on the outputs of teams and firms (Böheim et al., 2014; Trax et al., 2015) to evaluate the outcome of direct interactions, but potentially neglecting spillover effects through local labor market mechanisms (Kemeny and Cooke, 2018).

[1]It also includes the study of diversity as an outcome of economic phenomena.

Building on this general picture, this chapter contributes to this literature by estimating the impact of birthplace diversity and migration on wages within skill bins in local US labor markets. By defining diversity at this intermediate level, it attempts to disentangle the effect of diverse migrations from a general heterogeneity of skills related to national specializations (D'Amuri and Peri, 2014) to investigate the role of local complementarities and interactions. To do so, it studies the influence of diversity on individual wages either at the level of metropolitan occupations in their respective sectors or at the level of experience-education skill groups. The first level is likely to emphasize the mechanisms associated with the exchange of ideas or communication issues, as these are more likely to occur among individuals in the same local industries and professions due to the higher level of interactions and the local market mechanisms. Doing so, this research assesses whether the net effects of diversity go beyond a complementarity of occupations and/or general skills. The second definition builds on the idea that segments of American workers tend to change their occupations in response to migrations (Peri and Sparber, 2011; Peri, 2012). As immigrants and natives can self-select into occupations and industries, but are less flexible in the short run regarding their levels of education and professional experience, this stable definition allows for capturing complementarities between similar workers related to professional mobility and migration-induced specializations (Ottaviano and Peri, 2012). Endogeneity concerns are addressed through a shift-share instrumental strategy based on immigrant diasporas and placebo regressions following the procedure of Docquier et al. (2020).

As main results, I find a robust but slight positive link between workers' hourly wages and immigrant diversity in sector-occupational groups: A standard deviation increase of the metropolitan immigrant diversity of workers belonging to the same industry and occupation is associated with an increase of 0.689% of their hourly wages when controlling for metropolitan and local sector-occupational agglomeration. The estimated effect seems solely driven by workers without a college education. Defining worker groups as education-experience bins allowing for occupational complementarity and mobility to be captured, results highlight a positive impact of these groups' immigrant birthplace diversity on their native wages for low-skilled workers and a negative (insignificant) impact for high-skilled workers. These results suggest that even highly similar workers are still imperfect substitutes. The overall estimated impact is magnified with migrants originating from high-income countries that are either historically (proxied by genetics) strongly close or strongly distant to the US Population (*versus* intermediate distance).

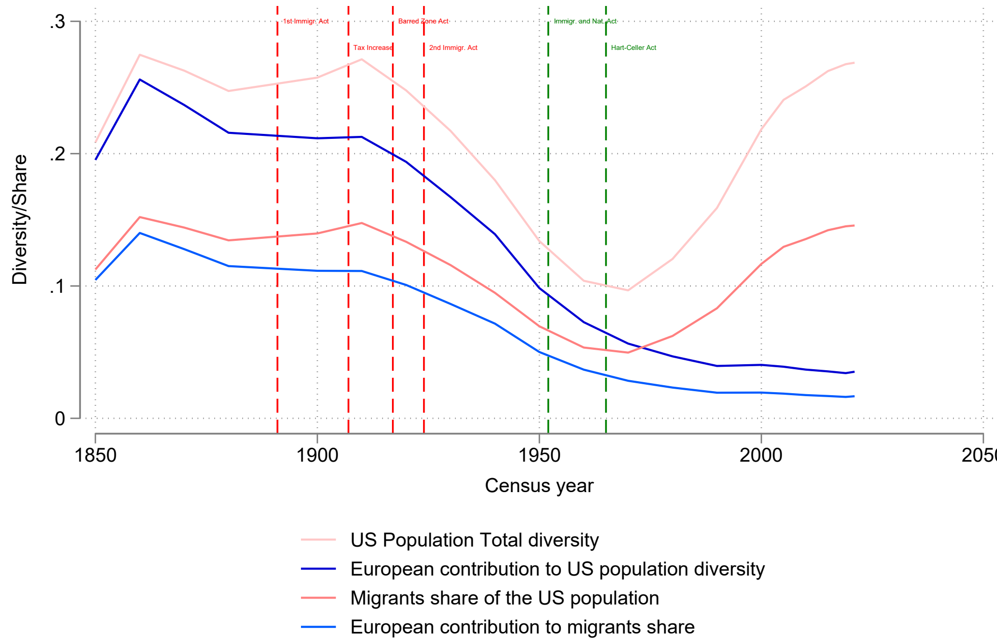
These results support two main conclusions: Firstly, that birthplace diversity decreases income inequality as it positively affects blue-collar workers' productivity, potentially through migrants' different sorting in occupations and industries than natives. Diversity also attenuates competition at highly local levels, potentially reflecting sorting in tasks within the same jobs, but this attenuation is not strongly robust and seems less effective with historically distant birthplaces, suggesting the presence of communication rigidities. The estimated net positive effects remain slight at any level considered.

This chapter is organized as follows: The second section briefly describes the migration context of the US, both historically and under the period covered by the present study. The third section offers a summary of the mechanisms considered in the economic literature and its main findings. A fourth section describes the data and the methodology used in this study, and a fifth section presents the main analysis results.

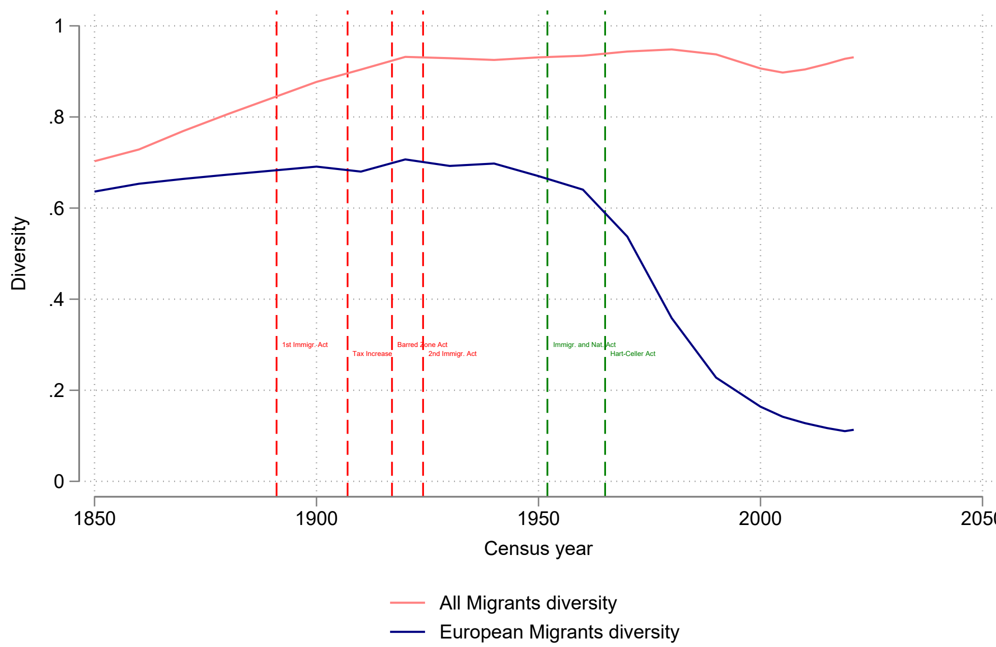
2.2 Migrations in the US: A brief history

The beginning of the 21st century offers an interesting context that contrasts with previous US historical migration history often studied: The fig. 2.1 (a) pictures the evolution of diversity measured as the probability of drawing two persons from two different birthplaces among the US population and migrant shares and the European contribution to these measures. It attests that most of the early past strong migration flows to the US were driven by European migrants. A highly documented sharp decrease in migrants and population diversity followed the successive introduction of anti-migration policies during the first half of the 20th century, including a national-origins quota system[2]. After its abandonment, birthplace diversity and the foreign share of the US population increased back to levels similar to those before the wave of anti-migration policies. It also resulted in a reduction in the European contribution to migration, while the diversity of European migrants fell sharply and overall migrant diversity remained stable in response to the end of quotas (fig. 2.1 (b)). This indicates an increase in the extra-European migration size and diversity, mostly driven by Asian and Central American origins (section 2.B.1 fig. 2.B7), that peaked in the most recent years. Such a variety of cultures, languages, and values that are more distant from those of native Americans is a relatively new form of birthplace diversity that has not yet been studied in depth.

[2]The Barred Zone Act of 1917. Another strong barrier to migration was initiated by the second Immigration Act of 1925.



(a)



(b)

Note: Diversity is defined here as the probability of drawing two persons from different birthplaces among the US, the migrants, or the European migrants population. The red hatched lines mark the introduction of major restrictive migration policies, and the green lines mark the removal of major barriers to migration. Source: Author's computations on the 2000-2021 ACS, 5% State 1970-1990 Census and 1% Census 1850-1960 IPUMS databases (Ruggles et al., 2023).

Figure 2.1: Long-run evolution of US birthplace diversity, migrant share (a) and migrants diversity (b)

2.3 The literature

2.3.1 Mechanisms

The gathering of people originating from a diverse set of countries is considered to reflect a wider range of local skills, norms, and ideas that, in turn, affect economic outcomes. Beyond indirect effects on productivity and growth, Alesina and Ferrara (2005) argue that a direct impact emerges from complementarities among workers with distinct cultural or demographic characteristics. Hong and Page (2001) establish a formal link between group composition and problem-solving ability, emphasizing the role of background diversity in achieving superior outcomes thanks to the presence of different paths of interpretation and algorithms. Workers born in different countries may not be perfect substitutes even within similar skill and experience groups, owing to differences in cultural backgrounds and educational experiences. Berliant and Fujita (2008) argue that it exacerbates the cross-fertilization of ideas at the firm level. The net effects of productivity may be the result of several opposing mechanisms, as adverse effects may also arise from communication barriers (Richard et al., 2002; Dale-Olsen and Finseraas, 2020), conflicts, uncooperative behavior, and weaker social ties and trust (Lazear, 1999; Alesina and La Ferrara, 2002). These elements support the existence of an optimal level of diversity, in line with the influential work of Ashraf and Galor (2013).

The seminal model of Ottaviano and Peri (2006) pictures the localized impact of population diversity in the presence of both supply and demand-side externalities: in their framework, the worker complementarities and learning externalities have the potential to enhance labor productivity, while a preference for population diversity may also attract workers. Florida (2002) argues that diverse and tolerant cities may draw in creative individuals, thereby fostering high-tech industries and research. Such attractiveness may also arise from the availability of different consumption varieties associated with a diverse population[3].

[3]Examining the validity of this second assertion in Germany, Vossen et al. (2019) find that diversity serves as a poor predictor of the presence of these workers. Instead, they appear to be attracted by traditional

The section 2.B.5 model extends Ottaviano and Peri (2006) to take into account this potential heterogeneity of preferences and productivity with respect to diversity, illustrating the potential influence of diversity on wages through these two channels.

2.3.2 Net effect of regional diversity

The magnitude and sign of the estimated net effect of birthplace diversity vary according to the level and scale of the research, as emphasized in the detailed reviews made by Kemeny (2017) and Ozgen (2021). Comparing countries, Alesina et al. (2016) found a positive and significant impact of birthplace diversity on national output per capita. Their results suggest that the gains from diversity arose from first-generation immigrants. They also echo the findings of Ashraf and Galor (2013) supporting a quadratic function linking development and genetic diversity. These non-monotonic relationships, implying an optimal level of diversity, are usually not supported locally for birthplace diversity: For instance, Böheim et al. (2014) find evidence of a fourth-order polynomial relationship between birthplace diversity and wages in Austrian firms, but its shape remains relatively monotonic. Most of the other studies at the subnational level do not find evidence of a quadratic relationship and keep a standard log-linear setting.

Due to the mechanisms described in the previous section, most of them take place at the regional or city levels to capture workers' interactions and local market mechanisms. Studying the influence of birthplace diversity on wages at the US State level, Docquier et al. (2020) find positive estimates for college-educated workers and insignificant positive estimates for the others. They also find an overall positive net impact at the commuting zone level. In Germany, Suedekum et al. (2014) find support for a positive but weak relationship for *high-skilled* immigrant nationality diversity at the regional level, and a negative relationship for *low-skilled* workers diversity. Using similar German data, Trax et al. (2015) provide evidence of a general positive effect both at the regional and firm levels. Limonov

amenities such as schools and other public goods.

and Nesena (2016) achieve the opposite conclusion for all workers in Russian regions with strong negative estimates. Ottaviano and Peri (2006) provides evidence of a positive impact of metropolitan birthplace diversity on native wages using American data between 1970 and 1990 and restricting their sample to the white 40 to 50-year-old workers. Focusing on the effect of ethnic diversity on native wages at the district level in the UK, Longhi (2013) fails to detect a significant relationship when correcting partially for individual unobserved heterogeneity. Using a sample of European western regions in 1991 and 2001, Bellini et al. (2013) found a positive relationship between NUTS3 nationality diversity and GDP per capita. The exact magnitude of the estimates from the closest specification from the present study can be appreciated in table 2.1: Without extreme estimates, it ranges from -0.94 to 0.55, with mostly weak positive values.

Article	indep.	dep.	scale	level	start	end	OLS	IV
Docquier et al. (2020)	Migr. div.	HS month. wage	USA	States	1960	2010		0.553*
Ottaviano and Peri (2006)	Migr. div.	annual wage	USA	Cities	1970	1990	0.14*	
Kemeny and Cooke (2018)	Migr. div.	annual wage	USA	Cities	1991	2008	0.403***	
Suedekum et al. (2014)	Migr. div.	Native daily wage	DE	Nuts3 Reg.	1995	2006	0.0325***	
Haus-Reve et al. (2021)	Migr. div.	annual wage	NO	Reg.	2001	2011	0.065***	0.048***
Limonov and Nesena (2016)	Migr. div.	month. wage	RU	Reg. (Adm)	2002	2010		-4.53**
Elias and Paradies (2016)	Tot. div.	weekly wage	AU	Reg. (LGA)	2001	2011	-0.012	0.203
Bakens et al. (2013)	Tot. div.	annual wage	NL	Cities	1999	2008	-0.51	-0.7
Kemeny and Cooke (2018)	Tot. div.	annual wage	USA	Cities	1991	2008	0.406***	0.432***
Nathan (2011)	Tot. div.	hourly wage	UK	CZ	1994	2008	0.322**	0.142
Charnysh (2019)	Past Migr. div.	personal income tax pc	PL	Munic.	1998	1998	0.24***	
Limonov and Nesena (2016)	Tot. div.	month. wage	RU	Reg. (Adm)	2002	2010	-2.34***	-0.94***
Dohse and Gold (2014)	Tot. div. (cont., Theil)	yearly Wage pc	Eur. (EU27)	Nuts2 Reg.	2005	2010	0.03***	
Maré and Poot (2019)	Tot. div.	annual wage	New Zealand	Cities	1981	2013	0.048***	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table displays OLS and IV estimates from the economic literature that explored the relationship between wages and diversity at the regional or city levels with a specification close to the present study. Indep. The column indep. indicates the measurement of diversity (Migr. div. if diversity is computed in the migrant population, Tot. div. if among the whole population. By default, it is a Herfindahl index, but Dohse and Gold (2014) used a Theil index instead. The dep. column indicates the measurement of wage (which group (HS is among college-educated workers) and over which period). Start and end columns indicate the period of the study.

Table 2.1: Selected sample of Regression Estimates - Diversity and wages

When considering cities and firm/workplace diversity impact on annual worker income in the US, Kemeny and Cooke (2018) found general positive effects, and they obtain positive results for specific groups of workers associated with complex problem-solving (Cooke and Kemeny, 2017). The estimated effects are bigger in cities with more inclusive institutions (Kemeny and Cooke, 2017). Overall, empirical evidence suggests a weak positive influence of migration diversity on wages. Most of the studies measure diversity using a Herfindahl, a Theil, or a Shannon index that imposes the same weight to each origin, regardless of their specific characteristic. The next section presents a few exceptions that attempted to take into account the different economic and cultural characteristics that birthplace diversity embodies into augmented indexes.

2.3.3 Economic and cultural heterogeneity

Few studies to date have attempted to integrate cultural and economic birthplace heterogeneity in the measurement of diversity. Notable exceptions are Alesina et al. (2016) at the country level, who found a quadratic relationship between the effect of birthplace diversity and cultural distance proxied by genetic distance, Tabellini (2020) who integrates religious and linguistic distances in a weighted migrant presence index, and Docquier et al. (2020) who obtained results suggesting that US gains from diversity at State level are maximized when it originates from the presence of immigrants either genetically or economically more distant from the natives, but not both. Evaluating the impact of diversity on well-being, Akay et al. (2017) find that the effect is magnified when the cultural distance is minimal.

2.4 Empirical framework

2.4.1 Data

To investigate the relationship between birthplace diversity and the workers' wages at local skill groups' levels, I use the Integrated Public Use Microdata Series[4] (IPUMS) that contains demographic and earnings information collected on the American population at the individual level by the U.S. Census Bureau. Specifically, I use the 2005, 2010, 2015, and 2019 American Community Survey (ACS) databases that together form independently pooled cross-sectional data. These data are weighted to adjust for the effects of nonresponse and to correct for survey undercoverage. IPUMS sample weighting allows the derivation of representative population estimates at aggregate levels. The geographic level of the analysis

[4]Steven Ruggles, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler and Matthew Sobek. IPUMS USA: Version 11.0 [dataset]. Minneapolis, MN: IPUMS, 2021. https://doi.org/10.18128/D010.V11.0

is the Metropolitan Statistical Area (MSA). They cover very populated areas and the adjacent communities with a strong economic and social link to the core, but their definition changes over time to reflect the growth of cities. For stability purposes, MSAs are defined using their 2013 areas from the U.S. Office of Management and Budget (OMB). Computing diversity at this level helps reduce the impact of spatial correlations found in smaller areas, as it proxies for local labor markets. Following Docquier et al. (2020), persons born in any of the 50 United States or American possessions[5] are considered natives, and the sample includes individuals aged 16 to 64. Hourly wages can be calculated based on annual work-related income and the number of weeks and hours per week spent at work, but the number of weeks per year is only available in 12-week intervals for the periods 2010 and 2015. I use the median between the two limits of the interval for them, following Ottaviano and Peri (2012). The main results are, however, robust to the exclusion of these periods. Descriptive statistics of the sample can be found in the section 2.B.1 table 2.B1. Genetic distances are from Spolaore and Wacziarg (2016) and GDP per capita from the World Bank.

2.4.2 Diversity measurement

This research focuses on immigrant birthplace diversity: Being born and raised in a different country than the other inhabitants of one's city results in a different cultural background, as well as different school training and problem-solving algorithms. This decision is in line with several authors who employ this form of diversity[6]. Most of this literature measures it using a fractionalization (Herfindahl) Index. If we define $s_{c,g}$ as the share of individuals from country g living in a region c, G the number of foreign birthplaces, and US as the native origin, the overall diversity index in an area c is:

[5]This includes American Samoa, U.S. Virgin Islands, Guam, and Puerto Rico.

[6]See Ottaviano and Peri (2006), Bellini et al. (2013), Kemeny (2017), Kemeny and Cooke (2018) or Docquier et al. (2020) among others.

$$Total\ diversity_c = \sum_{g}^{G+1} s_{cg}(1 - s_{cg}) = 2s_{c,US}(1 - s_{c,US}) + (1 - s_{c,US})^2 \sum_{g \neq US}^{G} s^M_{c,g}(1 - s^M_{c,g}) \quad (2.1)$$

Where $s^M_{c,g}$ is the share of migrants from country g in the foreign population and $s_{c,US}$ is the share of natives in region c. The equation shows that the index can be decomposed into a *between* and a *within* components: the left-hand side, called *between diversity*, is the component related to the presence of natives and migrants in the same geographical area. A major weakness of this index comes from this component, which is also related to economic outcomes, as higher wages will attract more migrants in general, resulting in more diversity. To avoid capturing this circular relationship, Alesina et al. (2016) typically compute the index for the migrant populations alone, controlling for migration incentives using $migrant\ share = 1 - s_{c,US}$. I followed them in using the same measurement. It represents the probability of drawing two individuals from two different birthplace groups in the immigrant population:

$$Migrant\ diversity_c = \sum_{g \neq US}^{G} s^M_{c,g}(1 - s^M_{c,g}) \quad (2.2)$$

2.4.3 Baseline equation and skill groups' definition

This research focuses on the effect of the migrant component of local skill groups' diversity on individual hourly wages. I use the following baseline empirical specification for the wage equation:

$$ln(wage_{i(cj)t}) = X'_{it}\phi + \alpha_1 div_{cjt} + \alpha_2 mig.\ share_{cjt} + \gamma_j + \gamma_c + \gamma_t + \varepsilon_{i(cj),t} \quad (2.3)$$

Where i indexes for the individual, c her MSA, j her skill group, and t the period. Local diversity div_{cjt} is the eq. (2.2) immigrant fractionalization index in a defined skill group

j. The dependent variable is computed as the hourly (constant 1999$) wages of individual native workers. This avoids capturing the changes in the total hours of work in the estimates, as with annual income. The term γ_j represents the skill group fixed effect. γ_c is an MSA fixed effect capturing the effect of geography and other fixed amenities. I add $X_{i,t}$, which is a vector of time-varying and unvarying individual characteristics, to control for individual heterogeneity in skill and preferences (it includes age, squared age, as well as race and gender dummies). Such controls form the main reasons behind the decision to define wages at individual levels: results may be sensitive to individual characteristics that are not taken into account with aggregated average wages (Longhi, 2013). The *mig. share* variable is the local share of immigrants. This term prevents the diversity coefficient from being closely tied to the proportion of foreign-born individuals. It therefore allows for disentangling diversity effects from the aggregate flow of immigration. This distinction is important as pictured in the section 2.B.1 fig. 2.B2 that shows a kernel-estimated function of the general relationship between metropolitan GDP per capita and total diversity, and its decomposition across immigrant diversity and general migration exposure components. Finally, $\varepsilon_{i(c,j),t}$ is an error term. I also control for MSA and group size in most specifications to avoid capturing pure agglomeration economies. This general equation forms the baseline for estimating the effect of birthplace diversity at both skill levels on wages.

Two skill groups' definitions are considered for the estimation: The first is the industrial occupational groups (1990 definition of industries and occupations). An example is the group of civil engineers working in the Electrical Machinery industry in Memphis. (The exhaustive list of occupations and industries is available in the section 2.B.4). Workers inside these groups are likely to be highly similar in skills. The second is the education and experience cells considered by Ottaviano and Peri (2012). They are defined by levels of education (No High School, High School, Some College, Bachelor, Master, PhD) and 5-years potential professional experience[7] bins. To the best of my knowledge, it is the first time that birthplace

[7]That is the number of years between the theoretical end of studies and current age.

diversity indices are computed at these local levels. The comparison between these groups has several advantages: estimation made using aggregated groups of workers provides the net effect of birthplace diversity, but it is complex to determine whether this does indeed capture the spillovers mentioned in the literature, or whether it is a reflection of skill diversity that may, in turn, be related to amenities or metropolitan industrial composition. Working at local levels is therefore expected to capture mostly the mechanisms of interest rather than the complementarity of occupations related to national specializations. On the other hand, as native workers are heterogeneous in terms of occupational mobility and as they may change their occupation in response to migration patterns, the sector-occupation level is likely to provide local estimates, as they will measure the relationship for the native workers who remain in (or directly moved to) their current occupation in response to its prevailing migration patterns. These are likely to be a specific subgroup of workers that faces a lack of occupational mobility. It may also partially capture the mobility of native higher-paid workers to other sectors and occupations in response to these patterns; the structure of the data does not permit a precise decomposition of the relative contributions of these two channels. It is for comparison that indices are computed at a more aggregated level of metropolitan experience-education skill cells, as it captures the occupational mobility. The estimates can therefore be considered less local.

2.4.4 Instrumental strategy

A circular causal mechanism is likely to characterise the link between diversity and both groups' wages: Alesina and Ferrara (2005) argue for an ambiguous sign bias arising from the fact that economically more affluent areas may attract migrants from all backgrounds (implying a positive bias) as well as migrants from specific countries (implying a negative bias). Additionally, endogeneity concerns partially stem from unobserved preferences[8] or

[8]e.g. for specific amenities related to diversity (Florida, 2002)

groups/individual characteristics[9]. Kemeny and Cooke (2018) argue that it should result in a positive bias. The micro approach of the analysis, which allows for controlling for the main individual, industrial, and occupational characteristics, may not be sufficient to prevent it.

To address these endogeneity concerns, this article employs a shift-share (Bartik) instrumental strategy. It has typically been used by Ottaviano and Peri (2006) in their specification for regional global wages and rents. The idea is that the growth rate of the migrant stock in c from country g, $m_{c,g}$, can be decomposed into a component specific to each birthplace and an idiosyncratic (both metropolitan and birthplace-specific) component:

$$m_{c,g} = m_g + \widetilde{m_{c,g}} \tag{2.4}$$

To get rid of the idiosyncratic component, the shift-share strategy consists in using predicted stocks to compute a predicted diversity index:

$$\hat{s}_{g,c,t} = (1 + m_{g,t-t_0}) s_{g,c,t_0} \tag{2.5}$$

where s_{g,c,t_0} is a defined past (at time t_0) stock of migrants from country g. In the case of this study, we consider migrant stocks in 2000. The use of the birthplace-specific component (that I compute using $m_{g,t-t_0} = (s_{g,t} - s_{g,t_0})/s_{g,t_0}$, the growth rate of the migrant stock from g among the US population) would allow retrieving a diversity index capturing a part of diversity related to the choice of moving in an area due to the past presence of people from the same origin[10]. The instrumental strategy relies on the fact that migrants tend to go to geographical areas where their communities are more present, but also to occupy the same types of positions in the same industries. First-stage estimates tend to confirm it is the case (section 2.B.1 table 2.B7). Several elements have to be addressed for such a strategy in general (see Goldsmith-Pinkham et al. (2020) and Borusyak et al. (2022) for

[9] e.g. specific skills locate according to the presence of specific industries (Combes et al., 2008) correlated with skills and affecting the migration patterns.

[10] See Beine et al. (2011) for empirical evidence of this phenomenon.

detailed discussions on this strategy): Firstly, the exclusion assumption may require that the initial stocks s_{g,c,t_0} are strictly exogenous to the future wages. Such an assumption is more likely to be met by taking stocks well before the periods of interest. A major weakness in the use of such an instrument with IPUMS data is that workers are only identified at the level of geographical entities such as PUMA or MSA for very recent periods (from 2000 onwards for the MSA using the IPUMS 5% Census); the farthest one can go back is 1960 when working at the level of States as Docquier et al. (2020) and Ottaviano and Peri (2006), which is a too broad geographic level for studying the previous mechanisms, or at the levels of counties that are likely to suffer from spatial endogeneity. Moreover, the past larger migrant communities participate more in the variations of the instrument. It means that one needs to be aware of specific characteristics related to these communities: If their specific characteristics influence wages in the long run, it may violate the exclusion restriction. The section 2.B.1 fig. 2.B7 shows the biggest contributor to US migrant population that may be problematic. Building on Borusyak et al. (2022), the shift-share remains a valid instrument even when these conditions are violated locally under the condition of the exogeneity of the aggregated (birthplace) US migrant flows. Under this assumption, one can carefully consider a causal interpretation of the results. The section 2.B.7 presents different tests of this assumption and replications excluding Mexicans.

2.5 Results

This section presents the empirical results investigating the influence of immigrant birthplace diversity on wages among local skill groups. It is structured as follows: The section 2.5.1 describes OLS and 2SLS estimates using the baseline specification, the section 2.5.2 exploits intranational migration as a placebo to explore the potential endogeneity with respect to income pull-factors, the section 2.5.3 explores the potential role of occupational sorting behind the findings, the section 2.5.4 shows estimates at various levels of analysis to assess the potential consequences of the previous findings and assess the influence of the level of analysis on the estimates, and the last section 2.5.5 allows for cultural and economic distances to be taken into account in the estimates.

2.5.1 Metropolitan skill group diversity analysis: Baseline OLS and 2SLS estimates

This section explores the relationship between wages and immigrant diversity computed at the metropolitan skill group levels. The table 2.2 presents the OLS and shift-share 2SLS estimates that allow for investigating the effect of birthplace diversity among immigrants in metropolitan industry-occupations, as well as in the metropolitan experience-education cells. Column 1 panel B 2SLS positive significant estimate implies that a standard deviation increase in a local industry-occupation immigrant diversity results in a 0.52% (exact variation) predicted wage increase for an American native in this occupation. Using the instrumental strategy shows little divergence from the OLS result. Turning to the education-experience cells diversity estimation in column 4, it reaches 1.39%, more than three times its OLS equivalent.

Intuitively, the mechanisms of task and algorithm complementarities would imply a gain of productivity from diversity in sectors or occupations where there is space for (or where outputs are) innovations. It is a common finding in research where a positive relationship is found between regional high-skilled diversity and wages, and low-skilled workers are associated with an absent or a negative influence (Docquier et al., 2020; Suedekum et al., 2014). For the US, Docquier et al. (2020) typically detect a positive association with the number of patents and high-skilled wages, but their estimates are insignificant for low-skilled wages. However, the estimates presented above do not seem to embody such mechanisms: When splitting the sample between college-educated and other workers, low-skilled workers' immigrant diversity estimates are significantly positive and highly similar between the sector-occupation groups in column 2 and the experience-education skill cells in column 5.

Even more surprising: they display opposing signs for high-skilled workers (columns 3 and 6), contrasting with findings at more aggregated levels. Additionally, the OLS and 2SLS estimates are close for Low-skilled workers, suggesting that the instrument nets out relatively little endogeneity for this group.

These asymmetric patterns across skill categories require careful interpretation. The sense of the bias in the OLS regression on wages is theoretically ambiguous, as income pull-factors can both attract diverse immigrants or trigger strong migration flows from a specific subsample of countries. This bias could differ across skill groups: Occupational group diversity in high-skilled occupations may reflect a political or employers' will to attract the very best among the world population for specific occupations (Miyagiwa, 1991; Kennedy, 2019), hence the insignificant results when (supposedly) achieving removal of income pull-factors in the estimates: The United States indeed attract workers in highly competitive high-skilled sectors and occupations, notably through favorable high-skilled migration policies (Kerr et al., 2015; Bound et al., 2017; Adovor et al., 2021). This will to attract the very best workers from all origins in high-skill sectors can induce a verticality among high-skill workers: diversity would reflect this will, which mechanically translates into higher wages, not because diversity influences productivity, but rather because the immigrants in diverse workers' populations are among the upper tail of their respective occupation. It implies a positive bias in the OLS. It would remain consistent with the opposite findings of Docquier et al. (2020) that suggest it is dominantly negative for high-skilled workers at regional levels[11], if they capture broader complementarities. On the other hand, low-skilled workers may be subject to horizontal complementarities: a craftsman may have origin-specific skills that are scarce among the other origins, for instance. However, low-skilled workers tend to be more sensitive to their networks for migration decisions (McKenzie and Rapoport, 2010; Beine et al., 2011), potentially in an endogenous manner, as their network may help to find a job

[11] Their equivalent estimate increases from 0.319 (OLS) to 0.372 for high-skilled diversity and high-skilled wages, and from 0.171 to (insignificant) -0.317 regarding low-skilled diversity and wages. Note that replicating their specification at the Metropolitan level using the current research data while controlling for individual characteristics (section 2.B.2 table 2.B3 col. 6 and 7) provides insignificant estimates close to 0.

in the destination region. In that case, the instrument could be unadapted for low-skilled workers: Even if the first stage regressions, available in the section 2.B.1 table 2.B7, and their Kleibergen and Paap F-statistics[12] indicate that shift-share predicted diversity has a strong explanatory power of actual diversity for industry-occupation and education-experience cells, both among low-skilled and high-skilled workers[13], it would still partially capture the income pull-factor if the diaspora is broadly spatially distributed. On the contrary, if the instrument is exogenous, the proximity of the OLS and 2SLS estimates instead suggests that the pull-factors bias is modest for low-skilled diversity. It would also imply that the benefits from diversity manifest differently at distinct levels across the high- and low-skilled workers. The section 2.5.2 investigates the presence of a pull-factor bias in the estimates of table 2.2 to evaluate the credibility of these last assertions, while the next paragraphs examine the robustness of these findings to heterogeneity and alternative income measurement.

Robustness checks

Alternative income variable - As the positive estimates may be rooted in occupational mobility and that workers may also receive other kinds of monetary transfer than wages, the section 2.B.1 table 2.B8 replicates the regressions using the quantile-defined occupational earning as the proportion of U.S. workers earning below the average annual income within the worker's current occupations, as an alternative dependent variable. As some workers are self-employed and may partially be remunerated through other forms of income than wages or tips, as the section 2.B.6 table 2.B12 different rankings suggest, it allows for testing if the education-experience skill cell migrant diversity is related to the relative overall occupational annual earnings. Interestingly, estimates are positive and significant for both college and non-college individuals (columns (1) to (3)), but they are positive and in-

[12]Note that the high F-statistics are not driven by the methodological choice of instrumenting both migrant share and diversity: When instrumenting only immigrant diversity as Docquier et al. (2020) in section 2.B.3 table 2.B5, they are higher. The estimates also remain similar for experience-education bins, and the industry-occupation high-skilled sample even reveals a negative coefficient.

[13]The strength of the shift-share instrument for high-skilled workers' migration patterns echoes the results of Docquier et al. (2020) and Mayda et al. (2022).

	Industry-occupation			Education-Experience Skill cells		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Low-skilled	High-skilled	All	Low-skilled	High-skilled
Panel A: OLS estimates						
Migrant diversity$_j$	0.0154***	0.0216***	0.0264***	0.0133	0.0216***	-0.0190**
	(4.34)	(4.33)	(4.86)	(1.54)	(3.22)	(-2.52)
Migrant share$_j$	-0.0395***	-0.0417***	-0.0186	0.0813***	0.0526**	0.243***
	(-2.67)	(-4.11)	(-1.19)	(3.24)	(2.44)	(5.22)
$ln(Density)$	0.0564*	0.0312	0.0652*	0.0607*	0.0585	0.0735*
	(1.73)	(0.84)	(1.84)	(1.66)	(1.49)	(1.84)
$ln(Density_j)$	0.00892***	0.00506***	0.0124***	0.0102*	-0.0170***	0.0101*
	(3.38)	(3.38)	(4.50)	(1.73)	(-3.65)	(1.82)
Panel B: IV estimates						
Migrant diversity$_j$	0.0199**	0.0211**	0.0107	0.0381*	0.0301*	-0.0231
	(2.34)	(2.25)	(0.90)	(1.76)	(1.92)	(-1.21)
Migrant share$_j$	-0.143***	-0.130***	-0.0985***	0.0581**	0.0241	0.366***
	(-5.35)	(-7.23)	(-2.71)	(2.59)	(1.04)	(6.00)
$ln(Density)$	0.0604*	0.0367	0.0818**	0.0641*	0.0584	0.0689*
	(1.70)	(0.96)	(2.14)	(1.72)	(1.47)	(1.66)
$ln(Density_j)$	0.0153***	0.00880***	0.0185***	0.00904	-0.0177***	0.0108
	(4.72)	(4.23)	(5.04)	(1.37)	(-3.82)	(1.62)
Time FE γ_t	Yes	Yes	Yes	Yes	Yes	Yes
Work. groups FE γ_j	Yes	Yes	Yes	Yes	Yes	Yes
MSA FE γ_c	Yes	Yes	Yes	Yes	Yes	Yes
Individual char. $X_{i,t}$	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,470,217	1,309,085	2,155,038	3,477,385	1,315,610	2,161,775
R-squared	0.409	0.307	0.402	0.317	0.210	0.283
KP F-Stat	336.4	288.4	360.5	249.4	269.6	287.9

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, t statistics in parentheses. Robust SE clustered at the MSA-level. Panel A displays OLS estimates based on eq. (2.3) by skill groups. Migrant diversity indices and shares are computed at the metropolitan worker skill cell level. Worker groups are either defined as industry-occupational groups or as experience-education skill cells. Panel B contains 2SLS estimates (Migrant share and diversity are instrumented using a shift-share strategy). The dependent variable is log of native \$1999 hourly wage. X_{it} is a vector of individual characteristics containing age, age squared, racial and gender dummies. Source: Author's elaboration on IPUMS 2000-2019 ACS data.

Table 2.2: OLS and 2SLS estimates by skill groups - workers groups immigrant diversity and wages

significant when regressing occupational earnings quantile in 1990 on contemporary migrant diversity (columns (4) to (6)). It suggests that the effect of skill cell diversity is primarily driven by shifts in occupational income rather than by changes in occupational composition among native workers. Nevertheless, it is important to interpret these results with caution, as the insignificance of the estimates may be attributable to the limited variation of the 1990 quantiles, and this variable is likely to capture changes in hours and income in the whole population.

Sectoral and cohort heterogeneity - I now seek to test whether the diversity estimates are heterogeneous across sectors and specific occupations. An interesting feature in this separation lies in the tradable or non-tradable characteristic of goods and services related to the sectors argued by Moretti (2004): For instance, service-oriented sectors can hardly export them to other cities, so they can be considered as non-tradable sectors facing regional prices. This translates into a need to locate production activities in the same geographical area as consumers. This makes it more difficult for firms to be mobile and react to local wage changes. Manufacturing sectors, on the other hand, can more easily export their production across the national territory, facing national prices; in case of local wage increases that are not linked to productivity increases, firms can therefore spatially relocate to regions with compensating differences. A focus on the tradable sectors would then allow for considering a relationship between immigrant diversity and productivity as an increase in wages. Additionally, it allows for assessing the external validity of the estimates across industries and occupations. Aggregated groups are built using the aggregate IPUMS industries as well as the occupational groups defined by David and Dorn (2013)[14]. The graphs in the section 2.B.1 fig. 2.B4 display estimates from the same regressions as table 2.2 columns (1) and (4) while discriminating by aggregated sector. They picture a relative homogeneity of the main findings across aggregated industries. Replicating the regressions within aggregated industry-occupation cells (section 2.B.1 fig. 2.B6), however, reveals heterogeneity

[14]The section 2.B.4 describes the occupations and industries inside each category.

across occupations within some sectors. Nevertheless, the estimates for the manufacturing (tradable) sector are particularly high and significant overall and for most of its occupational groups. When computing the 2SLS coefficient directly by aggregated occupational groups (section 2.B.1 fig. 2.B5), one can observe a relative heterogeneity in the estimates for *High-skilled* occupations, while the coefficients are similar among *low-skilled* occupations. As an additional robustness check, replications of the industry-occupation diversity regressions by age cohorts reveal highly homogeneous results (section 2.B.1 fig. 2.B3).

2.5.2 Placebo regressions: Exploiting intranational migration patterns

The previous section shows little divergence between IV and OLS estimates, especially for low-skilled workers. It may result from a poor performance of the Bartik instrument if it also captures the circular causality mentioned in section 2.4.4. This section uses internal migration patterns as a placebo in order to investigate the presence of this potential endogeneity threat and limit its influence. To do so, I create a State Herfindahl diversity index computed within the metropolitan skills group workers, following Docquier et al. (2020). This variable measures the probability of drawing two workers born in two different US states among the natives who were born in another US state. I also define $Native\ share_j$ as the share of metropolitan native workers in group j born in a US state other than where they currently live. Estimating eq. (2.3) with these measures of US internal migration can serve as a Placebo regression: the income pull-factor of migration should also trigger intranational migrations. Moreover, controlling for native diversity and migrations can then help to mitigate the potential bias. The OLS estimates are depicted in table 2.3.

Interestingly, the coefficients of native diversity are positive in columns 1 and 3, indicating that wages are higher in occupations attracting workers from various US states. Education-experience bins express the inverse pattern (table 2.3 Columns 5 and 7), suggesting that higher wages mostly attract high-skilled native migrants from specific locations at the experience-education skill level. These estimates suggest the presence of an income pull-factor correlated with diversity for college-educated workers[15], consistent with their high degree of mobility (Glaeser et al., 2004; Kennan and Walker, 2011), and, to a lesser extent, for low-skilled workers in industry-occupation skill groups that present lower estimates. While these placebo tests do not necessarily mean that low-skilled foreign immigrants' diversity is not affected by local labor market performance, it does not detect it for native migrations. It supports that the positive foreign immigrant diversity effect estimated (previously and in table 2.3 column 6) for this group reflects a genuine causal relationship rather than labor market sorting. The conclusion differs for college-educated workers, where the placebo results suggest that higher wages do attract more geographically diverse native workers. This pattern may extend to foreign immigrants and weaken identification, even with a shift-share instrumental strategy[16].

To mitigate this potential identification threat, table 2.3 pair columns present the OLS estimates associated with immigrant diversity while controlling for native migration patterns. It obliterates the effect of industry-occupational immigrant diversity (col. 2 and 4). Columns 6 and 8 show that the pattern previously detected regarding experience-education groups' migrant diversity remains unaffected by the inclusion of native migration covariates. These results require deep caution: Firstly, intranational migrations may act as bad controls if they are correlated with mechanisms induced by migrant diversity. Secondly, applying the shift-share instrumental strategy to this new specification in table 2.4 highlights different results: the low-skilled workers' sample has now slightly positive and significant industry-occupation diversity 2SLS estimates, while all high-skilled 2SLS estimates become insignificant, despite similar values.

[15] These results slightly contrast with estimates at the state level that tend to be positively high but insignificant (Docquier et al., 2020).

[16] The similarity of the OLS and 2SLS estimates may originate from a violation of the exclusion restriction in the high-skilled context. Instrumenting native by predicted foreign diversity in section 2.B.3 table 2.B10 shows insignificant low second-stage estimates and K-P F Wald statistic inferior to the standard thresholds despite significant augmented correlation.

	Industry-occupation				Education-Experience Skill cells			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Low-skilled	Low-skilled	High-skilled	High-skilled	Low-skilled	Low-skilled	High-skilled	High-skilled
Migrant diversity$_j$		0.00731		-0.00636		0.0199***		-0.0169**
		(1.46)		(-1.63)		(3.08)		(-2.33)
Migrant share$_j$		-0.0471***		-0.0172		0.0534**		0.228***
		(-4.56)		(-1.11)		(2.42)		(5.10)
Native diversity$_j$	0.0130***	0.00917**	0.0280***	0.0282***	-0.00937	-0.0102	-0.0614***	-0.0528***
	(3.08)	(2.11)	(5.33)	(5.12)	(-0.83)	(-0.89)	(-4.52)	(-4.40)
Native share$_j$	0.00159	0.00245	0.0297***	0.0298***	0.0122	0.00375	0.0747***	0.0574***
	(0.46)	(0.71)	(8.30)	(8.13)	(0.65)	(0.21)	(3.44)	(3.35)
$ln(density)$	0.0358	0.0363	0.0764**	0.0777**	0.0371	0.0339	0.0971***	0.0808**
	(1.07)	(1.08)	(2.33)	(2.37)	(1.01)	(0.93)	(2.63)	(2.17)
$ln(density_j)$	0.0114***	0.0130***	0.0213***	0.0226***	-0.0119*	-0.0158**	0.0168**	0.0169*
	(6.08)	(6.46)	(6.19)	(5.83)	(-1.78)	(-2.28)	(2.04)	(1.91)
Time FE γ_t	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Work. groups FE γ_j	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA FE γ_c	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual char. $X_{i,t}$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,309,085	1,309,085	2,155,038	2,155,038	1,315,610	1,315,610	2,161,775	2,161,775
R-squared	0.308	0.308	0.402	0.402	0.210	0.210	0.283	0.283

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, t statistics in parentheses. Robust SE clustered at the MSA-level. OLS estimates based on eq. (2.3) by skill groups. An intranational immigrant is defined here as a worker born in a U.S. state other than the U.S. state in which he lives. Intranational Migrant diversity indices and shares are computed at the metropolitan worker skill cell level. Worker groups are either defined as industry-occupational groups or as Ottaviano and Peri (2012) skill cells. The dependent variable is the log of the native \$1999 individual hourly wage. X_{it} is a vector of individual characteristics containing age, age squared, racial, and gender dummies. Source: Author's elaboration on IPUMS 2000-2019 ACS data.

Table 2.3: OLS estimates by skill groups - Intranational migration patterns and wages

	Industry-occupation		Education-Experience Skill cells	
	(1)	(2)	(3)	(4)
	Low-skilled	High-skilled	Low-skilled	High-skilled
Migrant diversity$_j$	0.0154*	0.00734	0.0224	-0.0130
	(1.76)	(0.72)	(1.28)	(-0.68)
Migrant share$_j$	-0.148***	-0.125***	-0.0764***	0.0614
	(-7.53)	(-3.24)	(-3.15)	(1.00)
Nat. diversity$_j$	0.00158	0.0234***	-0.00716	-0.0587***
	(0.30)	(3.13)	(-0.61)	(-4.38)
Nat. share$_j$	0.00487	0.0317***	0.0137	0.0753***
	(1.16)	(6.50)	(0.68)	(3.55)
1st stage estimates				
pr. migrant div.$_j$	0.610***	0.534***	0.425***	0.356***
	(32.92)	(37.97)	(27.91)	(26.84)
Density variables	Yes	Yes	Yes	Yes
γ_t	Yes	Yes	Yes	Yes
γ_c	Yes	Yes	Yes	Yes
γ_j	Yes	Yes	Yes	Yes
X_{it}	Yes	Yes	Yes	Yes
Observations	1,137,565	1,909,005	1,288,688	2,119,414
K-P F-Stat.	386.8	395.3	419.4	126.7

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, t statistics in parentheses. Robust SE clustered at the MSA-level. An intranational immigrant is defined here as a worker born in a U.S. state other than the U.S. state in which he lives. Intranational Migrant diversity indices and shares are computed at the metropolitan worker group level. Worker groups are either defined as industry-occupational groups or as Ottaviano and Peri (2012) experience-education skill cells. Migrant share and diversity are instrumented using a shift-share strategy, while native migration variables are not instrumented and serve as controls. The dependent variable is the log of the native $1999 individual hourly wage. Density variables include the (log of) metropolitan and metropolitan group j densities. X_{it} is a vector of individual characteristics containing age, age squared, racial, and gender dummies. Source: Author's elaboration on IPUMS 2000–2019 ACS data.

Table 2.4: 2SLS estimates by skill group - accounting for intranational migration patterns

2.5.3 Exploring occupational complementarities and mobility

The previous sections present evidence of a positive influence of immigrant diversity on local skill group wages, mostly driven by the low-skilled workers. This relationship suggests the presence of complementarities and mechanisms beyond the highly local complementarities within specific occupations[17]. But it also suggests positive effects among workers with similar education and experience. This section explores a potential mechanism underlying these results: occupational sorting. Peri and Sparber (2011) show, using panel data, that native workers in the United States tend to specialize in occupations with greater communication content in response to immigration shocks, whereas college-educated immigrants disproportionately sort into occupations requiring quantitative and analytical skills. Consistent with the patterns identified above, Gu and Sparber (2017) find that occupational adjustment in response to immigration is primarily concentrated among low-skilled workers in experience-education groups. If similar workers specialized in different occupations and tasks within their skill groups according to their respective origins, diversity would also reflect these complementarities. Additionally, a more diverse environment may increase the demand for communication skills for which natives may have a comparative advantage. I test for different sorting patterns between immigrants and natives associated with diversity using the relative content of tasks of an occupation computed from David and Dorn (2013), who provide indexes of occupational routine, abstract, and manual task contents. These measures are not perfect as it only embodies a fragment of tasks performed in an occupation, but as it is fixed and specific to each occupation, differences necessarily mean selection in different occupations. Unfortunately, in the absence of panel data, mobility is not directly observable, and I am unable to distinguish between direct occupational complementarities and adjustment through occupational reallocation. The estimate could, for instance, represent a tendency of migrants to sort into jobs in lower positions on the labor market than natives with similar qualifications (Zorlu, 2016). I adapt the specification of Peri and Sparber

[17]Note that it survives the inclusion of intranational migrations among the covariates for the education-experience bins only with OLS estimates, but remains quantitatively similar with the shift-share strategy. The coefficient for the industry-occupations becomes small and insignificant with those controls in the OLS case, and reappears high and significant in the 2SLS specification.

(2011) to the aggregated data[18]:

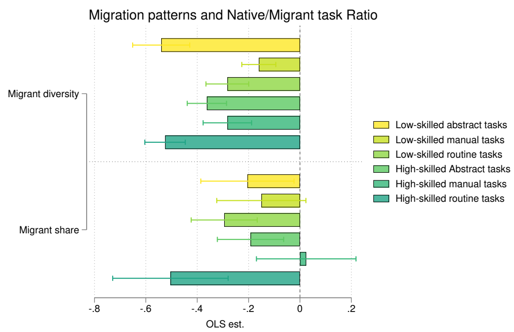
$$\frac{Tasks_{cjt,US}}{Tasks_{cjt,Migrant}} = \alpha + \alpha_1 div_{cjt} + \alpha_2 mig.\ share_{cjt} + \alpha_3 ln(density_{cjt}) + \gamma_j + \gamma_c + \gamma_t + \varepsilon_{cjt} \quad (2.6)$$

Where the dependent variable is the ratio of average task content (routine, abstract, or manual) at the experience-education level. The sample is split between high-skilled and low-skilled workers. The fig. 2.2 shows the estimates as well as 95% confidence intervals. OLS estimates almost all indicate a significant difference in task contents between native and migrant in skill cells more exposed to diversity and migration flows. When using the instrumental shift-share strategy to attenuate the role of endogenous pull factors, migrant flows are not associated with significant differences anymore, but diversity systematically is. These results show a sorting pattern likely to be rooted in direct occupational complementarities, but it also suggests that a greater variety of origins indeed embodies a variety of skills different from those of the natives and that migrants are likely to differ in skills. However, investigating whether diversity is associated with a sorting of natives toward more abstract-related functions relative to manual or routine tasks yields negligible results (section 2.B.3 table 2.B11).

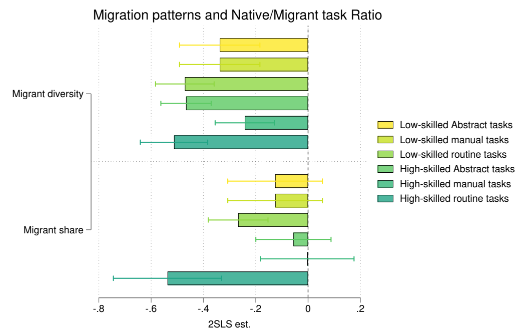
2.5.4 Analysis at aggregated and local levels

Previous results emphasize the different sorting of migrants and natives across occupations. The placebo regressions support internal migration patterns related to higher wages. Taken together, these results suggest the presence of both occupational and geographic mobilities (Ortega and Verdugo, 2022) of US natives that are difficult to capture in the net effect of diversity. Furthermore, diversity indices were computed at the education-experience cell or industry-occupational levels, while the wage was defined individually. It allowed the use of controls for individual heterogeneity. They were also conducted at the MSA geographic unit

[18] aggregated at the level of the metropolitan skill cell



(a)



(b)

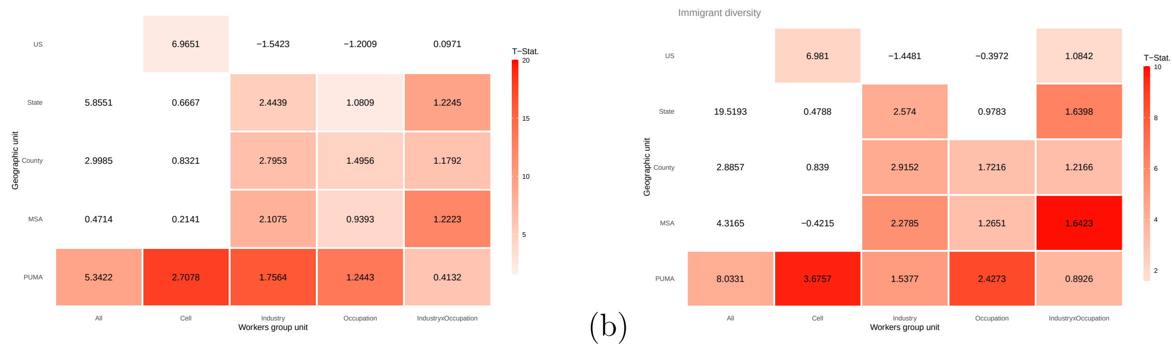
Note: The graph pictures OLS (a) or shift-share 2SLS (b) coefficient from a regression with the standardized (*z-score*) average experience-education bin (Ottaviano and Peri, 2012) ratio of task occupational content from David and Dorn (2013) between natives and migrants as a dependent variable. An observation is a skill cell at a defined year in a defined metropolis. The lower the dependent variable ratio, the more natives select in occupation less intensive in those tasks than immigrants. The independent variables are the skill cell immigrant birthplace index and share. Controls include skill cell, time and MSA fixed effects as well as populations in the local cell. Source: Author's elaboration on IPUMS 2000-2019 ACS and David and Dorn (2013)'s data.

Figure 2.2: Migration patterns and occupational sorting within experience-education cell by skill group, OLS (a) and 2SLS esimates (b)

level, which excludes a wide part of the US. This section replicates them using average wages defined at various levels, matching the levels of definition of the main independent variable. It allows the inclusion of potential mobility in the net estimated effects by investigating the implications for average wages in different groups of workers (All, experience-education cell, industry, occupation, industry-occupation) at various geographic levels (USA, State, County, MSA, PUMA[19]). This picture of the heterogeneous influence of birthplace diversity according to its level of definition follows three main purposes: Firstly, it captures (or does not) both the influence of workers' (im)mobility across areas, industries, and occupations: the wider the area of definition, the more the results capture the influence of geographically highly mobile workers. MSA-level only captures the influence of mobility from inside, but not the wage changes of ancient residents moving outside. The same reasoning applies to the workers' categories. Secondly, it allows the inclusion of rural areas and smaller cities in the analysis. Thirdly, it allows for comparing the estimates with most of the literature. The related (population-weighted) OLS and 2SLS estimates are depicted in the fig. 2.3 heatplot.

[19]The Public Use Microdata Area (PUMA) cover geographic areas with a population size of 100,000–200,000 within a state, created by the U.S. Census Bureau.

To facilitate the comparison between the estimates, the diversity indices are standardized. The cells are coloured according to the value of the *T-statistic* and are white if the coefficient is insignificant (10% threshold). The table 2.2 main results are replicated in fig. 2.3 fourth row, 2nd column for skill cells and fifth for industry-occupation groups; their comparison, as well as the general pattern emerging from all the estimates, lead to the same assessments: birthplace diversity is significantly and positively associated with hourly wages when exposure to diversity is at a rather local level (regional and lower than the cell levels). A notable exception is the significant 7.21% (exact var.) predicted US exp native wage increase associated with one more standard deviation in immigrant birthplace diversity. This level is particularly relevant from a macro perspective, as it captures both geographic and occupational mobility among native workers who are similar in terms of years of education and experience. It ensures that any complementarities within these groups are captured at the national level. It therefore suggests a general positive net impact of birthplace diversity in the US. This result may partially be driven by intranational migrations due to the insignificance at the state, county, and MSA levels, but it may also capture crude complementarities. On the other hand, no group based on industries and occupations is associated with significant results for the national area; the similarities between the results at the county and the MSA levels support that the inclusion of the rest of the US territory does not affect the main conclusions of the previous analysis. The lowest geographic level, PUMA, is systematically associated with a positive estimate.



(a) (b)

Note: Each squared cell represents an OLS (a) or shift-share 2SLS (b) coefficient from a regression of the (log of) average hourly \$1999 wage on a standardized (*z-score*) immigrant birthplace index, both computed at different levels and following the eq. (2.3) without the individual-level variables. They are multiplied by 100 to reflect approximated variations in %. The levels are defined geographically and according to workers' groups' definitions: "All" reflects the entire worker population in the area, Cell the Ottaviano and Peri (2012)'s experience-education skill cells, "Industry" the detailed 1990 industries listed in appendix C.2, "Occupation" the detailed 1990 occupations listed in appendix C.1 and "IndustryxOccupation" all the joint components of these two last groups' definitions. As areas, "US" is associated with the national level results, "State" the (US) states level, "County" the county level, "MSA" the Metropolitan statistical area level and "PUMA" the Public Use Microdata Area level, that is the smallest geographic unit available. The estimates are weighted according to the population inside these groups to retrieve results representative of the US. Darker cells represent higher *t-statistics*, and white cells are insignificant at the 10% threshold. Source: Author's elaboration on IPUMS 2000-2019 ACS data.

Figure 2.3: Estimated impact (in %) of a standard deviation birthplace diversity increase on predicted wages at aggregated levels - OLS (a) and 2SLS (b) estimates

2.5.5 Taking cultural and economic heterogeneity into account

The previous section emphasizes positive estimates at various local skill levels. The local levels of analysis used in this research are supposed to give more weight to interaction and less to complementarities, while still integrating local labor market mechanisms. The industry-occupation level is highly local and is supposed to capture interactions and complementarities among strongly similar workers. Education-experience bins give more weight to complementarities among less similar workers and occupational sorting. Considering origins embody skills, making diversity a vector of positive externalities and complementarities, suppose each origin also has a different contribution to the diversity net effect according to economic and cultural characteristics, notably because of stronger complementarities or coordination problems. This section attempts to incorporate such elements by taking economic and cultural distances into account. To do so, I follow Alesina et al. (2016) and Docquier et al. (2020), who use augmented indices that allow for heterogeneity in cultural and economic backgrounds to participate in the measurement of immigrant diversity. A weakness of the birthplace diversity index used so far lies in the fact that it considers birthplaces as equally distant groups (for instance, a city with 50% Canadian and 50% Belgian populations is equivalent in terms of diversity to a city with 50% Canadian and 50% Chinese populations). This hypothesis is likely to be violated if there are cultural or economic similarities between the USA and specific countries. As in the extension of Alesina et al. (2016), the use of augmented indices that take into account the cultural and economic differences between origins allows for relaxing this hypothesis:

$$Augm.\ Migrant\ diversity\ _c = \sum_{g \neq US} s^M_{c,g}(1 - s^M_{c,g}) d_{g,US} e_{g,US} \tag{2.7}$$

where $d_{g,US}$ is a measure of cultural distance between a reference group (US) and the group g, and e captures economic distance. To retrieve these weights, Alesina et al. (2016) use the following logistic function of a distance that is also used for the purpose of this research:

$$d_{US,g} = \frac{2}{1 + e^{-\theta_D (D_{US,g})}} \tag{2.8}$$

Where $D_{US,g}$ is a standardized measure of distance between the birthplace g and the natives. This article uses the (standardized) genetic distance as $D_{US,g}$ in the footsteps of Alesina et al. (2016). It measures historical connection[20] between populations. θ_D is a weighting parameter that can be used to increase or decrease the impact of $D_{US,g}$ in the Index (The higher the θ_D, the higher the weight of $D_{US,g}$ and the closer the augmented and the unaugmented index[21]). Following Alesina et al. (2016) and Docquier et al. (2020), indices are computed for each weight value within a range of 1 to 10. Economic distance weights are obtained following the same procedure; as the US GDP should be endogenous to

[20] In terms of temporal distance from the period the two populations were one single population.

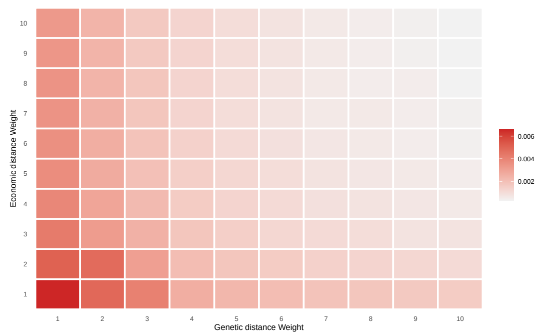
[21] Actually, the unaugmented index is the augmented index with $\theta \to +\infty$, which corresponds to a case where each birthplace receives a cultural distance from the natives that is equivalent and equal to 2.

US wages, the logistic transformation is applied to the standardized GDP per capita from the origin country Y_g instead:

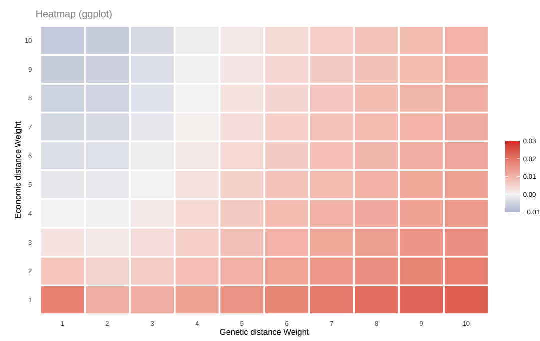
$$e_{US,g} = \frac{2}{1 + e^{\theta_E Y_g}} \tag{2.9}$$

These economic and genetic distances are used to retrieve 100 different diversity indices capturing cultural and economic distances to the US with different weights. The fig. 2.4 heatplots display estimates using the native sample at the industry-occupation (a) and skill cell (b) levels. Economic proximity is measured by the (standardized temporal average of) GDP per capita of the country of origin, as the difference with the US GDP is likely to be endogenous to wages. As a consequence, the bigger the weight parameter, the bigger the weight to low-income origins. The cultural distance is proxied by standardized genetic distance that reflects the historical connection between populations and captures cultural differences in a wide range of dimensions (Becker et al., 2020). The bigger the weight parameter, the bigger the weight to genetically distant countries origins.

The heatplot (a) suggests a maximized positive effect of a form of diversity resulting from migrations from countries richer and historically close to the US population at the metropolitan industry-occupation level. The size of the estimates strongly decreases, and they become insignificant when more weight is allocated to both low-income and/or genetically distant birthplaces. Replicating the heatplot using the migrant sample in section 2.B.1 fig. 2.B8 and computing distances to the representative migrant is associated to similar results for the cultural distance. The heatplot (b) reveals a similar feature that may reflect the phenomenon detected in (a), but another optimal corner is located at a form of diversity resulting from migrations from countries richer and historically distant to the US population at the metropolitan skill-cell level. Estimates also strongly decrease with more weight to low-income origins. Industry-occupation cells are smaller and are likely to be characterised by a stronger level of interactions than skill cells, while changes in wages in skill cells also capture changes in occupations. This difference between the two groups may arise from the stronger interactions present inside an industry-occupation group that may stress the importance of communication, while skill complementarities to the natives that cultural distance captures would also be detected at the broader skill cell level. As a consequence, the two corner optima in fig. 2.4 (b) may be respectively driven by skill cells characterised by a high capacity of occupational mobility and by skill cells with low occupational mobility.



(a)



(b)

Note: Each squared cell represents an OLS coefficient from a regression of the (log of) individual hourly \$1999 wage on a standardized (*z-score*) Augmented indices computed at the metropolitan sector-occupation pairs (a) and at the Ottaviano and Peri (2012) experience-education skill cells (b) levels. It follows eq. (2.3) specification while adding metropolitan total and workers group populations controls. Darker cells represent higher estimates, red and blue cells represent respectively positive and negative estimates. Y-axis represents Economic distance weights: the higher the parameter, the higher the weight to origins with a lower constant 2019\$ origin 2000-2019 temporal average GDP per capita. X-axis represents cultural distance weights: The higher the parameter, the higher the weight to origins with higher genetic distances from the US population. Source: Author's elaboration on IPUMS 2005-2019 ACS, World Bank, and Spolaore and Wacziarg (2016) data.

Figure 2.4: Augmented indices heatplots - Industry-occupation groups (a) and Education-Experience Skill cells (b)

2.6 Conclusion

This study investigates the relationship between immigrant birthplace diversity and wages in U.S. cities at the onset of the 21st century using a shift-share instrumental strategy and placebo regressions to address potential biases arising from unobserved preferences in migration patterns and income pull-factors. It differs from most of the related literature by focusing on diversity within metropolitan skill groups instead of regions or firms. The regional level tends to give more weight to broad complementarities, while the firm level emphasizes local interactions. The skill cell levels used in this chapter serve as intermediate levels that enable investigation of effects emphasizing broader interactions and more local complementarities at the level of metropolitan labor markets. By defining skill cells as industry-occupation groups, it emphasizes interactions among the same professions, while it attempts to capture metropolitan skill complementarities and occupational sorting effects by considering migrant diversity among workers similar in education and professional experience.

While different empirical findings in the migration economics literature support a positive effect of birthplace diversity on wages for college-educated workers and null or negative effects for non-college-educated workers (Docquier et al., 2020; Suedekum et al., 2014), the present study finds an inverse pattern: local diversity in education-experience and industry-occupation groups among non-college-educated workers is associated with wage gains when the overall findings are statistically insignificant for college-educated workers, with notable heterogeneity across occupations. These findings highlight different impacts of diversity and migration across college-educated and non-college-educated workers according to the levels of group aggregation definition. These particularities of low-skilled workers echo recent findings suggesting stronger occupational sorting among these workers in response to migration shocks (Gu and Sparber, 2017). Sorting phenomena are indeed supported by the data.

Finally, leveraging the composition of new migration flows using cultural and economic distance measurements, this article provides evidence that the benefits of diversity at the highly local occupation level in U.S. local industries are maximized when resulting from migrants born in high-income countries with historical and cultural links to the US population. On the other hand, it also provides evidence of a magnified impact at the wider education-experience bin level when diversity is the result of migration from high-income countries historically and culturally strongly distant from the US population[22]. These results align with the view that communication channels matter more for disaggregated workers' groups with many potential interactions, while skill complementarities play a more substantial role at broader levels of aggregation.

[22] *Versus* migration from moderately distant countries.

2.A Bibliography

Adovor, E., Czaika, M., Docquier, F., Moullan, Y., 2021. Medical brain drain: how many, where and why? Journal of health economics 76, 102409.

Ager, P., Brückner, M., 2013. Cultural diversity and economic growth: Evidence from the US during the age of mass migration. European Economic Review 64, 76–97.

Akay, A., Constant, A., Giulietti, C., Guzi, M., 2017. Ethnic diversity and well-being. Journal of Population Economics 30, 265–306.

Alesina, A., Ferrara, E.L., 2005. Ethnic diversity and economic performance. Journal of economic literature 43, 762–800.

Alesina, A., Harnoss, J., Rapoport, H., 2016. Birthplace diversity and economic prosperity. Journal of Economic Growth 21, 101–138.

Alesina, A., La Ferrara, E., 2002. Who trusts others? Journal of public economics 85, 207–234.

Alguacil, M., Alamá-Sabater, L., 2021. Migration in spain: The role of cultural diversity revisited. Politics and Governance 9, 118–132.

Ashraf, Q., Galor, O., 2013. The "Out of africa" hypothesis, human genetic diversity, and comparative economic development. American Economic Review 103, 1–46.

Autor, D., Dorn, D., Hanson, G.H., 2013. The china syndrome: Local labor market effects of import competition in the united states. American economic review 103, 2121–2168.

Aysheshim, K., Hinson, J.R., Panek, S.D., 2020. A primer on local area gross domestic product methodology. Survey of Current Business 100, 1–13.

Bakens, J., Mulder, P., Nijkamp, P., 2013. Economic impacts of cultural diversity in the netherlands: Productivity, utility, and sorting. Journal of Regional Science 53, 8–36.

Becker, A., Enke, B., Falk, A., 2020. Ancient origins of the global variation in economic preferences, in: AEA Papers and Proceedings, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203. pp. 319–323.

Beine, M., Docquier, F., Özden, Ç., 2011. Diasporas. Journal of Development Economics 95, 30–41.

Bellini, E., Ottaviano, G.I., Pinelli, D., Prarolo, G., 2013. Cultural diversity and economic performance: evidence from european regions, in: Geography, institutions and regional economic performance. Springer, pp. 121–141.

Berliant, M., Fujita, M., 2008. Knowledge creation as a square dance on the hilbert cube. International Economic Review 49, 1251–1295.

Böheim, R., Horvath, T., Mayr, K., 2014. Birthplace Diversity and Productivity Spill-Overs in Firms. CESifo Working Paper 5025. Munich. URL: https://hdl.handle.net/10419/103133.

Borusyak, K., Hull, P., Jaravel, X., 2022. Quasi-experimental shift-share research designs. The Review of Economic Studies 89, 181–213.

Bound, J., Khanna, G., Morales, N., 2017. Understanding the Economic Impact of the H-1B Program on the US. Technical Report. National Bureau of Economic Research.

Brunow, S., Brenzel, H., 2011. The effect of a culturally diverse population on regional income in EU regions. Technical Report. WIFO Working Papers.

Charnysh, V., 2019. Diversity, institutions, and economic outcomes: Post-wwii displacement in poland. American Political Science Review 113, 423–441.

Clark, T.N., Lloyd, R., Wong, K.K., Jain, P., 2002. Amenities drive urban growth. Journal of urban affairs 24, 493–515.

Combes, P.P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: Sorting matters! Journal of urban economics 63, 723–742.

Cooke, A., Kemeny, T., 2017. Cities, immigrant diversity, and complex problem solving. Research Policy 46, 1175–1185.

Dale-Olsen, H., Finseraas, H., 2020. Linguistic diversity and workplace productivity. Labour Economics 64, 101813.

David, H., Dorn, D., 2013. The growth of low-skill service jobs and the polarization of the us labor market. American economic review 103, 1553–97.

Di Berardino, C., D'Ingiullo, D., Furia, D., Cartone, A., 2021. Immigration diversity and regional economic growth. Economia Politica , 1–24.

D'Ippoliti, C., 2011. Economics and diversity. Routledge.

Docquier, F., Turati, R., Valette, J., Vasilakis, C., 2020. Birthplace diversity and economic growth: evidence from the us states in the post-world war ii period. Journal of Economic Geography 20, 321–354.

Dohse, D., Gold, R., 2014. Determining the impact of cultural diversity on regional economies in Europe. Technical Report. WWWforEurope working paper.

D'Amuri, F., Peri, G., 2014. Immigration, jobs, and employment protection: evidence from europe before and during the great recession. Journal of the European Economic Association 12, 432–464.

Elias, A., Paradies, Y., 2016. The regional impact of cultural diversity on wages: evidence from australia. IZA Journal of Migration 5, 1–24.

Florida, R., 2002. The economic geography of talent. Annals of the Association of American geographers 92, 743–755.

Florida, R., Mellander, C., Stolarick, K., 2008. Inside the black box of regional development—human capital, the creative class and tolerance. Journal of economic geography 8, 615–649.

Glaeser, E.L., Saiz, A., Burtless, G., Strange, W.C., 2004. The rise of the skilled city. Brookings-Wharton papers on urban affairs , 47–105.

Goldsmith-Pinkham, P., Sorkin, I., Swift, H., 2020. Bartik instruments: What, when, why, and how. American Economic Review 110, 2586–2624.

Gu, E., Sparber, C., 2017. The native-born occupational skill response to immigration within education and experience cells. Eastern Economic Journal 43, 426–450.

Haus-Reve, S., Cooke, A., Fitjar, R.D., Kemeny, T., 2021. Does assimilation shape the economic value of immigrant diversity? Economic Geography 97, 117–139.

Hong, L., Page, S.E., 2001. Problem solving by heterogeneous agents. Journal of economic theory 97, 123–163.

Kemeny, T., 2017. Immigrant diversity and economic performance in cities. International Regional Science Review 40, 164–208.

Kemeny, T., Cooke, A., 2017. Urban immigrant diversity and inclusive institutions. Economic Geography 93, 267–291.

Kemeny, T., Cooke, A., 2018. Spillovers from immigrant diversity in cities. Journal of Economic Geography 18, 213–245.

Kennan, J., Walker, J.R., 2011. The effect of expected income on individual migration decisions. Econometrica 79, 211–251.

Kennedy, A., 2019. The politics of skilled immigration: Explaining the ups and downs of the US h-1b visa program. International Migration Review 53, 346–370.

Kerr, S.P., Kerr, W.R., Lincoln, W.F., 2015. Skilled immigration and the employment structures of US firms. Journal of Labor Economics 33, S147–S186.

Kets, W., Sandroni, A., 2021. A theory of strategic uncertainty and cultural diversity. The Review of Economic Studies 88, 287–333.

Lazear, E.P., 1999. Globalisation and the market for team-mates. The Economic Journal 109, 15–40.

Limonov, L., Nesena, M., 2016. Regional cultural diversity in russia: does it matter for regional economic performance? Area development and policy 1, 63–93.

Longhi, S., 2013. Impact of cultural diversity on wages, evidence from panel data. Regional Science and Urban Economics 43, 797–807.

Maré, D.C., Poot, J., 2019. Valuing cultural diversity of cities. Available at SSRN 3477056 .

Mayda, A.M., Orefice, G., Santoni, G., 2022. Skilled immigration, task allocation and the innovation of firms. CESifo Working Paper .

McKenzie, D., Rapoport, H., 2010. Self-selection patterns in Mexico-US migration: the role of migration networks. the Review of Economics and Statistics 92, 811–821.

Minale, L., Rocha, R., Vigna, B., 2024. Immigrant Diversity and Long-Run Development. Technical Report. RF Berlin-CReAM Discussion Paper Series.

Miyagiwa, K., 1991. Scale economies in education and the brain drain problem. International economic review , 743–759.

Moretti, E., 2004. Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. Journal of econometrics 121, 175–212.

Nakao, K., Treas, J., 1994. Updating occupational prestige and socioeconomic scores: How the new measures measure up. Sociological methodology , 1–72.

Nam, C.B., Boyd, M., 2004. Occupational status in 2000; over a century of census-based measurement. Population Research and Policy Review 23, 327–358.

Nathan, M., 2011. The long term impacts of migration in british cities: Diversity, wages, employment and prices .

Nijkamp, P., Poot, J., Bakens, J., 2015. The economics of cultural diversity. Edward Elgar Publishing.

Ortega, J., Verdugo, G., 2022. Who stays and who leaves? immigration and the selection of natives across locations. Journal of Economic Geography 22, 221–260.

Ottaviano, G.I., Peri, G., 2006. The economic value of cultural diversity: evidence from us cities. Journal of Economic geography 6, 9–44.

Ottaviano, G.I., Peri, G., 2012. Rethinking the effect of immigration on wages. Journal of the European economic association 10, 152–197.

Ozgen, C., 2021. The economics of diversity: Innovation, productivity and the labour market. Journal of Economic Surveys 35, 1168–1216.

Ozgen, C., Nijkamp, P., Poot, J., 2017. The elusive effects of workplace diversity on innovation. Papers in Regional Science 96, S29–S49.

Parrotta, P., Pozzoli, D., Pytlikova, M., 2014. The nexus between labor diversity and firm's innovation. Journal of Population Economics 27, 303–364.

Peri, G., 2012. The effect of immigration on productivity: Evidence from us states. Review of Economics and Statistics 94, 348–358.

Peri, G., Sparber, C., 2011. Highly educated immigrants and native occupational choice. Industrial Relations: A journal of economy and society 50, 385–411.

Polachek, S.W., Chiswick, C., Rapoport, H., 2006. The economics of immigration and social diversity. Emerald Group Publishing Limited.

Richard, O.C., Kochan, T.A., McMillan-Capehart, A., 2002. The impact of visible diversity on organizational effectiveness: Disclosing the contents in pandora's black box. Journal of Business & Management 8, 265–265.

Rodríguez-Pose, A., Von Berlepsch, V., 2019. Does population diversity matter for economic development in the very long term? historic migration, diversity and county wealth in the us. European Journal of Population 35, 873–911.

Roupakias, S., Dimou, S., 2020. Impact of cultural diversity on local labor markets. evidence from Greece's "age of mass migration". The Manchester School 88, 282–304.

Ruggles, S., Flood, S., Sobek, M., Brockman, D., Cooper, G., Richards, S., Schouweiler, M., 2023. IPUMS USA: Version 13.0 [dataset]. Minneapolis, MN: IPUMS. URL: `https://doi.org/10.18128/D010.V13.0`, doi:`10.18128/D010.V13.0`. version 13.0.

Spolaore, E., Wacziarg, R., 2016. Ancestry, language and culture, in: The Palgrave handbook of economics and language. Springer, pp. 174–211.

Stirling, A., 1998. On the economics and analysis of diversity. Science Policy Research Unit (SPRU), Electronic Working Papers Series, Paper 28, 1–156.

Suedekum, J., Wolf, K., Blien, U., 2014. Cultural diversity and local labour markets. Regional Studies 48, 173–191.

Tabellini, M., 2020. Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration. The Review of Economic Studies 87, 454–486.

Trax, M., Brunow, S., Suedekum, J., 2015. Cultural diversity and plant-level productivity. Regional Science and Urban Economics 53, 85–96.

Vossen, D., Sternberg, R., Alfken, C., 2019. Internal migration of the 'creative class' in Germany. Regional Studies 53, 1359–1370.

Zorlu, A., 2016. Immigrants' occupational mobility—Down and back up again. IZA World of Labor. Institute for the Study of Labor (IZA).

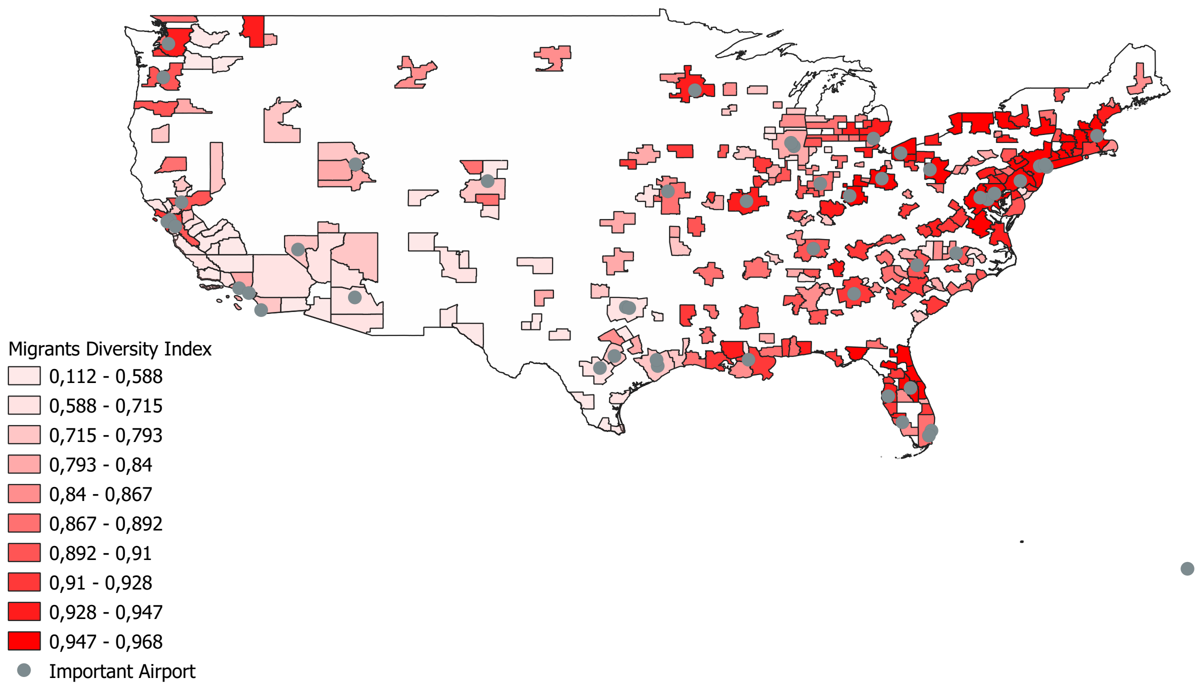
2.B Appendix

2.B.1 Descriptive analysis appendix

Variables	Mean	SD	Min	25% Q.	Median	75% Q.	Max
Individual variables							
wage	19.7223	172.3385	0.0007	8.1731	13.4615	21.6346	336,000
Migrant Share	0.1878	0.1260	0.0061	0.0796	0.1622	0.2809	0.4943
Migr. diversity Index	0.8548	0.1411	0.0839	0.7897	0.9075	0.9557	0.9758
Age	39.1503	13.6462	16	28	39	50	64
Racial, Gender and Education Dummies							
College-educated	0.5016	0.5000	0	0	1	1	1
Amerindian	0.0135	0.1153	0	0	0	0	1
Asian	0.0644	0.2455	0	0	0	0	1
Black	0.1257	0.3315	0	0	0	0	1
Pacific	0.0035	0.0587	0	0	0	0	1
White	0.7503	0.4328	0	1	1	1	1
Other race	0.0694	0.2542	0	0	0	0	1
Female	0.5131	0.4998	0	0	1	1	1
Metropolitan variables							
Descr. statistics computed on 2005-2019 averages at yearly metropolitan units (1,308 year x MSA)							
$\overline{wage}_c$	15.6493	3.0949	10.2750	13.6660	15.0823	16.9006	47.6724
GDP per capita	35.4530	10.0528	15.4321	28.6436	34.4291	40.2511	125.1175
Migrant Share	0.1237	0.0954	0.0061	0.0571	0.0937	0.1587	0.4943
Migr. diversity Index	0.8013	0.1821	0.0839	0.7456	0.8700	0.9248	0.9758
Density	7.1623	7.2722	0.1689	2.8850	5.1086	8.8485	60.8752
Coll.-ed. share	0.4976	0.0819	0.2866	0.4392	0.5002	0.5500	0.7608
Female Share	0.5036	0.0134	0.4148	0.4970	0.5053	0.5115	0.5391
Metropolitan Sector-occupational variables							
Descr. statistics computed on 2005-2019 averages at yearly metropolitan-sector-occupational groups (315,767 groups) using groups with at least one migrants.							
$\overline{wage}_j$	18.7415	79.9163	0.0008	8.7769	13.6369	21.3154	44,249.38
Migrant Share$_j$	0.5849	0.3591	0.0014	0.2426	0.5379	1	1
Migr. diversity Index$_j$	0.1508	0.2622	0.0000	0.0000	0.0000	0.3259	0.9692
Share$_j$	0.0010	0.0027	0.0000	0.0001	0.0002	0.0008	0.2584
Population$_j$	1,071.0893	3,612.2562	2	126	307	833	280,144
pred. Migr. diversity$_j$	0.0438	0.1213	0.0000	0.0000	0.0000	0.0000	0.8877
Metropolitan Education-Experience cell variables							
Descr. statistics computed on 2005-2019 averages at yearly metropolitan Ottaviano and Peri (2012) skill groups (14,618 groups).							
$\overline{wage}_j$	19.1956	126.3364	0.08271	10.5843	14.8335	22.1577	29,499.58
Migrant Share$_j$	0.1506	0.1927	0	0.0094	0.0828	0.2046	1
Migr. diversity Index$_j$	0.3995	0.3623	0.0000	0.0000	0.4552	0.7509	0.9733
Share$_j$	0.0118	0.0097	0.0000	0.0041	0.0104	0.0175	0.2010
Population$_j$	13,969.05	6,4824.93	4	1,229	3,263	9,150	2,444,835
pred. Migr. diversity$_j$	0.2651	0.3457	0.0000	0.0000	0.0000	0.5956	0.9724

Note: The data are from the US IPUMS ACS 2005-2010-2015-2019 databases. Descriptive statistics for Metropolitan variables are computed on the sample of Metropolitan Statistical Areas and the sample of Metropolitan workers groups for the Metropolitan Sector-occupational variables. Monetary variables are expressed in constant (1999) \$. $\overline{X}$ refers to the mean of variable X within a defined (metropolitan or metropolitan-sector-occupational) group.

Table 2.B1: Descriptive statistics

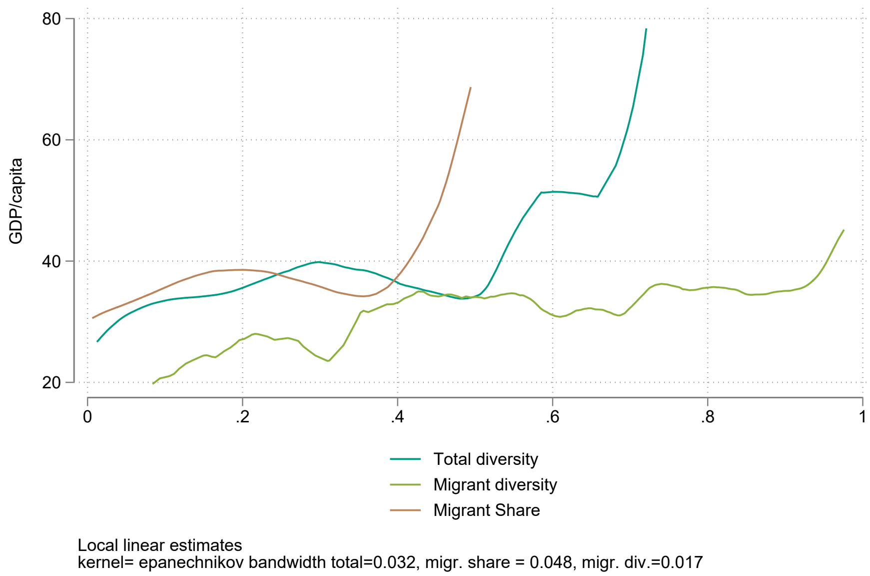


Note: The map shows the temporal mean of the birthplace diversity index computed on of the migrant population across metropolitan statistical areas, as well as the US airports handling more than 3 million passengers annually. Author's computation on IPUMS 2000, 2005, 2010, 2015 and 2019 ACS census and federal Aviation Administration Passenger Boarding Database of 2016. Alaska is not represented.

Figure 2.B1: Spatial distribution of birthplace migrant diversity, temporal mean 2000-2019

	Migr. div.$_j$	pr. Migr. div.$_j$	Total div.$_j$	Migr. share$_j$	Share$_j$	Density$_j$
Migr. diversity$_j$	1.0000					
pred. Migr. diversity$_j$	0.6233	1.0000				
Total diversity$_j$	0.6487	0.4144	1.0000			
Migr. share$_j$	0.2325	0.1594	0.4038	1.0000		
Share$_j$	0.1962	0.1510	0.1162	-0.0162	1.0000	
Density$_j$	0.3876	0.2791	0.2186	0.0285	0.6670	1.0000
Population$_j$	0.4338	0.3015	0.2392	0.0485	0.3561	0.7344

Table 2.B2: Diversity indexes and Migrants shares correlation matrix, metropolitan sector-occupational level



Note: Kernel regressions computed using optimal bandwidths. Source: Author's computation on 2019 ACS and Bureau of Economic Analysis 2019 GDP (Aysheshim et al., 2020) databases.

Figure 2.B2: Kernel regressions: Metropolitan GDP per capita (1999$) as a function of metropolitan migrants share in the labor force or total and migrant diversity indexes

2.B.2 Metropolitan diversity analysis

This section explores the relationship between wages and immigrant diversity computed at the metropolitan (MSA) level. The table 2.B3 provides replications of the estimates from the literature[23] but defining diversity at the US metropolitan level, considering native hourly wages, accounting for individual heterogeneity, and applying a weighting for the metropolitan population. The first column with a time fixed effects and the second, with sectoral-occupational fixed effects, show estimates close to the usual results when working with the mean wage at the local level. However, the estimates shrink and become insignificant when adding MSA fixed effects. These results hold when discriminating for college education in columns (4) and (5) and when computing diversity within these skill groups (columns (6) and (7)) while controlling for their size. It contrasts with similar studies carried out over similar timeframes and measuring salaries on a more aggregated basis (annual or monthly). The table 2.B4 also provides results for the migrants sample and for models allowing the

[23]The section 2.B.6 table 2.1 lists literature OLS and IV estimates following a similar methodology.

estimates to vary according to the metropolitan population. The results are similar but suggest that low-skilled workers in highly populated areas tend to have lower wages when exposed to higher levels of migration in general and when the migration is more diversified. Overall, these results suggest an influence close to a null effect.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	All	All	Low-Skill	High-Skill	Low-Skill div.	High-Skill div.
Migr. diversity	0.517***	0.256***	-0.00836	0.0178	-0.0232	-0.00789	-0.00686
	(8.73)	(4.47)	(-0.27)	(0.47)	(-0.63)	(-0.35)	(-0.16)
Migrant share	0.904***	0.547***	0.299***	0.165	0.321***	0.0225	0.461***
	(6.65)	(3.74)	(3.09)	(1.57)	(2.83)	(0.29)	(3.26)
$ln(Density)$		0.0178**	0.0536	0.0351	0.0748**	0.0229	0.111**
		(2.28)	(1.62)	(0.94)	(2.11)	(0.36)	(2.36)
$ln(Density_{LS\ or\ HS})$						0.0210	-0.0715**
						(0.48)	(-2.27)
Time FE γ_t	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA FE γ_c	No	No	Yes	Yes	Yes	Yes	Yes
Work. groups FE γ_j	No	Yes	Yes	Yes	Yes	Yes	Yes
Individual char. X_{it}	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,477,385	3,470,217	3,470,217	1,309,085	2,155,038	1,309,085	2,155,038
R-squared	0.0193	0.403	0.409	0.307	0.402	0.307	0.402

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, t statistics in parentheses. Robust SE clustered at the MSA-level. The dependent variable is (log of) $1999 constant hourly wage. Migrant diversity and share are computed at the metropolitan level (and among metropolitan education groups (College-educated (High-skilled) and the other workers (low-skilled)) in columns (6) and (7)). Individual characteristics include age, age squared, race, and gender dummies. $ln(Density_{LS\ or\ HS})$ is the metropolitan density of low-skilled workers in regression (6) and of High-skilled workers in regression (7). Source: Author's elaboration on IPUMS 2005-2019 ACS data.

Table 2.B3: OLS estimates - Metropolitan Migrant birthplace diversity and native hourly wages

	All		Migrants		Natives	
	(1)	(2)	(3)	(4)	(5)	(6)
	Low-skilled	High-skilled	Low-skilled	High-skilled	Low-skilled	High-skilled
Panel A: Baseline Model						
Migr. diversity	0.00320	-0.0406	-0.0287	-0.0710	0.0190	-0.0236
	(0.08)	(-0.94)	(-0.34)	(-0.64)	(0.51)	(-0.64)
Migrant share	-0.139	0.217^{*}	-0.381^{***}	0.0737	0.175^{*}	0.340^{***}
	(-1.38)	(1.75)	(-2.72)	(0.39)	(1.70)	(3.10)
$\overline{ln(Density)}$	0.0349	0.0572	0.0396	-0.0624	0.0323	0.0746^{**}
	(1.03)	(1.63)	(0.69)	(-1.02)	(0.93)	(2.18)
Panel B: Density interaction						
Migr. diversity	-0.129^{**}	-0.0669	-0.117	-0.0590	-0.130^{**}	-0.0659
	(-2.14)	(-1.31)	(-1.10)	(-0.54)	(-2.06)	(-1.43)
x $\overline{ln(Density)}$	-0.139^{***}	-0.0274	-0.136^{**}	0.0324	-0.151^{***}	-0.0443
	(-3.40)	(-0.78)	(-1.99)	(0.39)	(-3.48)	(-1.41)
Migrant share	-0.171^{*}	0.220^{*}	-0.420^{***}	0.127	0.131	0.333^{***}
	(-1.66)	(1.81)	(-3.16)	(0.60)	(1.23)	(2.97)
x $\overline{ln(Density)}$	-0.157^{**}	-0.0661	-0.0187	-0.0675	-0.165^{**}	-0.0667
	(-2.05)	(-0.74)	(-0.21)	(-0.41)	(-1.99)	(-0.87)
$\overline{ln(Density)}$	0.190^{***}	0.0944^{*}	0.153^{*}	-0.0751	0.199^{***}	0.127^{***}
	(3.67)	(1.96)	(1.83)	(-0.77)	(3.56)	(2.70)
γ_t	Yes	Yes	Yes	Yes	Yes	Yes
γ_j	Yes	Yes	Yes	Yes	Yes	Yes
γ_c	Yes	Yes	Yes	Yes	Yes	Yes
$X_{i,t}$	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,656,731	2,567,590	342,666	407,327	1,309,085	2,155,038
R-squared	0.282	0.398	0.214	0.411	0.307	0.402

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Baseline Model Panel is a reproduction of table 2.B3 regressions by skill and origin groups while density interaction panel also includes density interaction terms. t statistics in parentheses. Robust se clustered at MSA-level. Dependent variable: hourly nominal wage, \$1999. $\overline{ln(Density)}$ is demeaned (log of) MSA population density. x $\overline{ln(Density)}$ is the interaction between $\overline{ln(Density)}$ and the above variable. X_{it} is the vector of individual characteristics including age, age squared, college-education, racial and gender dummies. Fixed Effects γ indexes: i=worker, j=industry-occupation, c=MSA, t=year. Source: Author's elaboration on IPUMS 2005-2019 ACS data.

Table 2.B4: OLS estimates by skill and origin groups - Metropolitan Migrant birthplace diversity and native hourly wages

2.B.3 Regression analysis appendix

	Industry-occupation			Education-Experience Skill cells		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Low-skilled	High-skilled	All	Low-skilled	High-skilled
Migrant diversity$_j$	-0.0112	-0.00191	-0.0175*	0.0325	0.0269	-0.0235
	(-1.45)	(-0.23)	(-1.79)	(1.49)	(1.45)	(-1.20)
Migrant share$_j$	-0.0368**	-0.0447***	-0.0129	0.0769***	0.0534**	0.244***
	(-2.50)	(-4.39)	(-0.77)	(2.89)	(2.44)	(5.01)
$ln(Density)$	0.0591*	0.0378	0.0787**	0.0634*	0.0320	0.0757**
	(1.91)	(1.13)	(2.33)	(1.90)	(0.89)	(1.99)
$ln(Density_j)$	0.0242***	0.0151***	0.0279***	0.0158*	-0.0168**	0.0168**
	(7.18)	(7.02)	(7.53)	(1.92)	(-2.35)	(1.97)
Time FE γ_t	Yes	Yes	Yes	Yes	Yes	Yes
Work. groups FE γ_j	Yes	Yes	Yes	Yes	Yes	Yes
MSA FE γ_c	Yes	Yes	Yes	Yes	Yes	Yes
Individual char. $X_{i,t}$	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,470,217	1,309,085	2,155,038	3,477,385	1,315,610	2,161,775
KP F-Stat	712.2	675.3	703.5	484.3	673.0	437.1

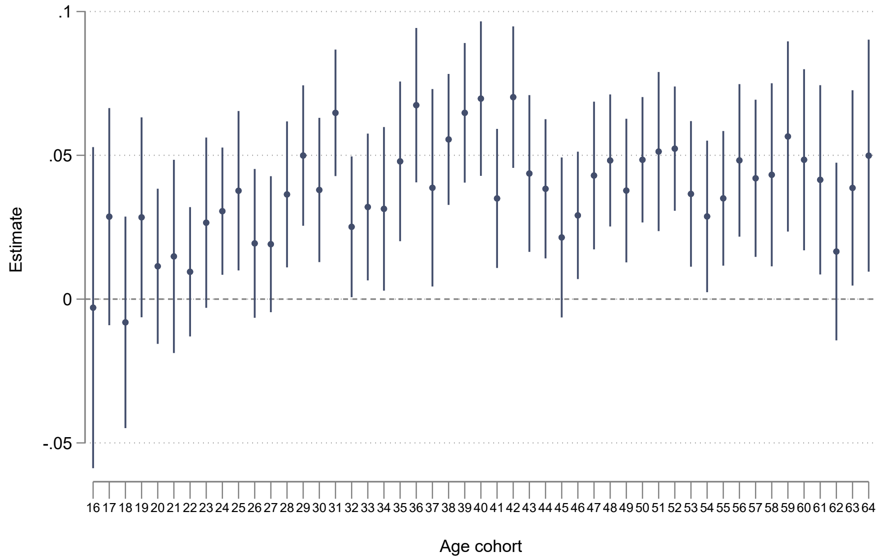
Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, t statistics in parentheses. Robust SE clustered at the MSA-level. Migrant diversity indices and shares are computed at the metropolitan worker skill cell level. Workers groups are either defined as industry-occupational groups or as Ottaviano and Peri (2012) skill cells. Dependent variable is log of native \$1999 hourly wage. X_{it} is a vector of individual characteristics containing age, age squared, racial and gender dummies. Source: Author's elaboration on IPUMS 2005–2019 ACS data.

Table 2.B5: 2SLS Estimates by skill groups, one single instrument - workers groups immigrant diversity and wages

	Industry-occupation			Education-Experience Skill cells		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Low-skilled	High-skilled	All	Low-skilled	High-skilled
Panel A: Low migration exposure sample						
Migrant diversity$_j$	0.00693	0.00369	0.00431	0.00620	0.0116*	-0.0136*
	(1.55)	(0.59)	(0.78)	(1.10)	(1.94)	(-1.96)
Migrant share$_j$	0.0256	0.0287	0.0199	0.126***	0.0918***	0.224***
	(1.61)	(1.57)	(0.90)	(4.46)	(3.62)	(4.98)
$ln(Density)$	0.0486	0.0291	0.0707*	0.0549	0.0370	0.0743*
	(1.44)	(0.78)	(1.96)	(1.40)	(0.92)	(1.70)
$ln(Density_j)$	0.00722***	0.00476***	0.00880***	0.00365	-0.0131***	0.00339
	(2.69)	(2.85)	(2.74)	(0.99)	(-2.99)	(0.88)
Observations	2.797.116	1.041.267	1.749.746	2.804.326	1.047.831	1.756.495
R-squared	0.407	0.319	0.397	0.315	0.218	0.280
Panel B: High migration exposure sample						
Migrant diversity$_j$	0.0296***	0.0253*	0.0283**	0.0184	0.0661***	-0.0564**
	(2.86)	(1.91)	(2.41)	(0.76)	(2.93)	(-2.24)
Migrant share$_j$	-0.0570***	-0.0760***	-0.0260	0.132***	0.103***	0.228***
	(-3.57)	(-4.60)	(-1.45)	(3.26)	(3.02)	(3.71)
$ln(Density)$	0.0915	0.0955	0.0947	0.151*	0.176**	0.157*
	(1.46)	(1.64)	(1.38)	(1.88)	(2.52)	(1.73)
$ln(Density_j)$	0.0217***	0.0102***	0.0296***	0.0138	-0.0253**	0.00648
	(7.01)	(3.91)	(7.96)	(1.33)	(-2.36)	(0.60)
Observations	670,402	265,468	402,895	673,058	267,778	405,278
R-squared	0.429	0.282	0.430	0.332	0.191	0.298
Time FE γ_t	Yes	Yes	Yes	Yes	Yes	Yes
Work. groups FE γ_j	Yes	Yes	Yes	Yes	Yes	Yes
MSA FE γ_c	Yes	Yes	Yes	Yes	Yes	Yes
Individual char. $X_{i,t}$	Yes	Yes	Yes	Yes	Yes	Yes

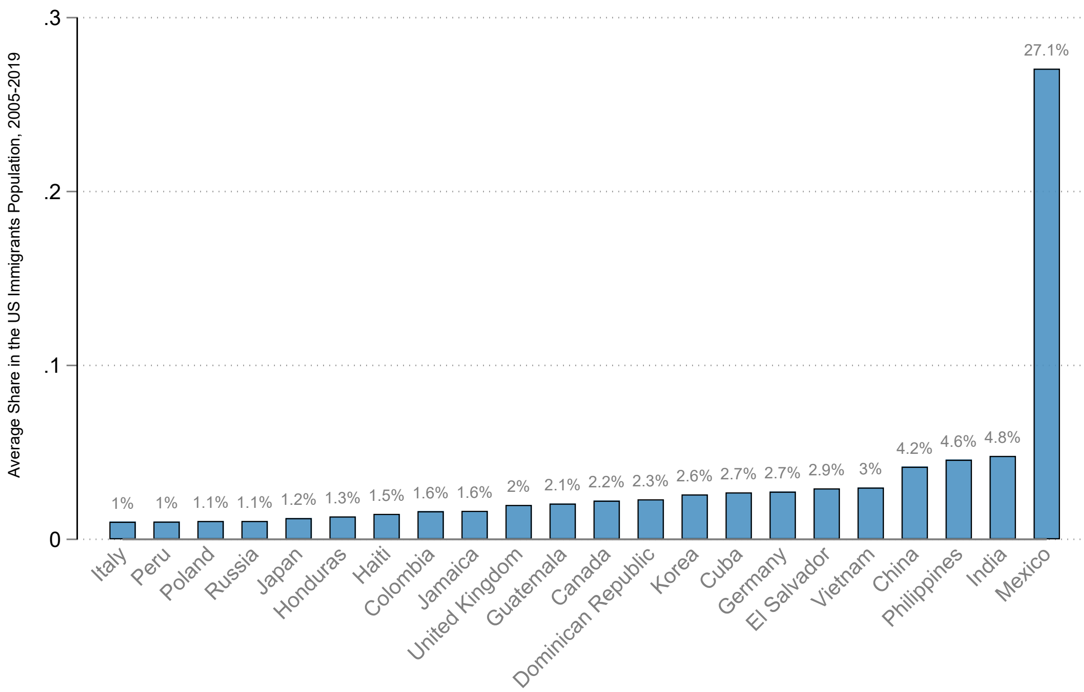
Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, t statistics in parentheses. Robust se clustered at MSA-level. Low and high migration exposure Panels are a replication of table 2.2 Panel A regressions using the sample of workers groups respectively under or above the median industry-occupation migration exposure (23%). Migrant diversity indexes and shares are computed at metropolitan worker groups level. Workers groups are either defined as industry-occupational groups or as Ottaviano and Peri (2012) skill cells. Dependent variable is log of native \$1999 hourly wage. X_{it} is a vector of individual characteristics containing age, age squared, racial and gender dummies. Source: Author's elaboration on IPUMS 2005-2019 ACS data.

Table 2.B6: OLS and 2SLS estimates by skill groups and migration exposure - workers groups migrant diversity



Note: OLS industry-occupation birthplace diversity estimates (eq. (2.3) with the addition of metropolitan and local workers group populations controls) by age cohort. Source: Author's elaboration on IPUMS 2005-2019 ACS data. Robust se clustered at the MSA-level. The dependent variable is (log of) $1999 constant hourly wage.

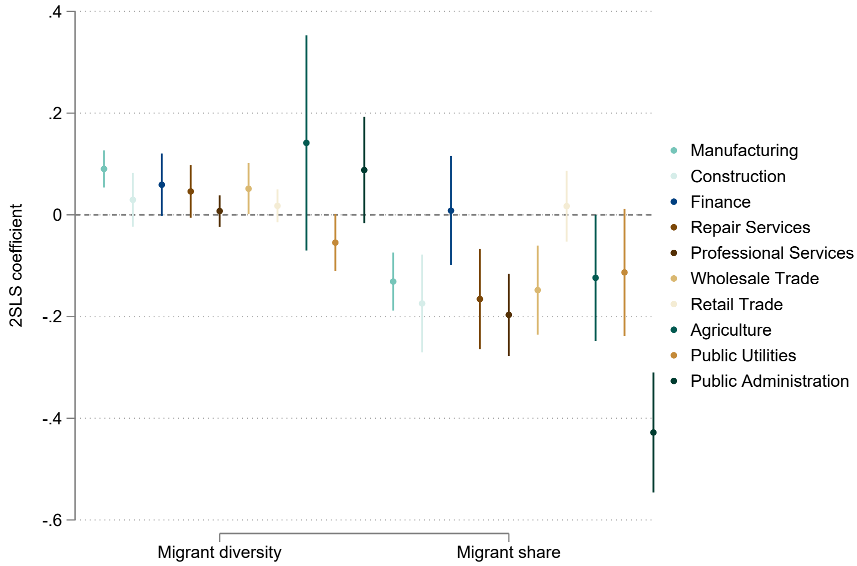
Figure 2.B3: OLS Industry-occupation diversity immigrant estimates by Age Cohort



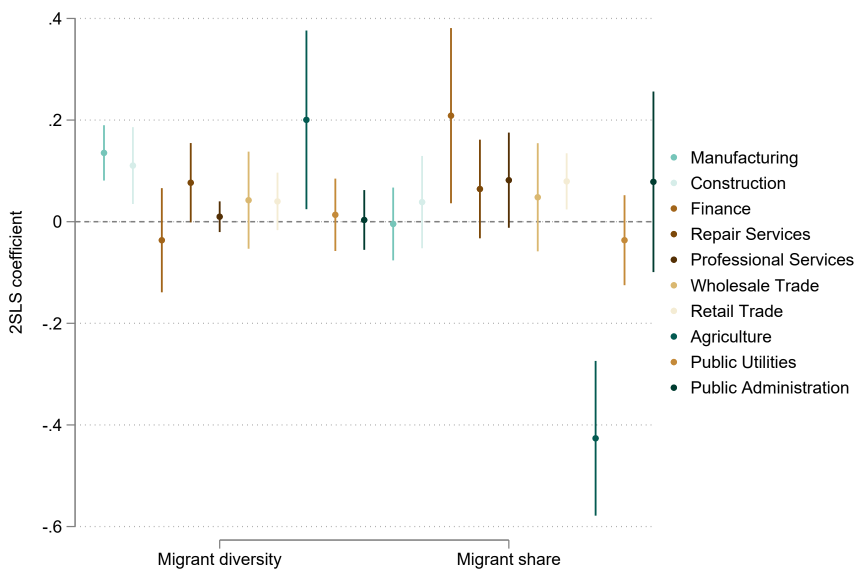
Note: The figure orders the 22 bigger US immigrant birthplace communities 2005-2019 by their average share among the total US immigrant population. Source: Author's computation on 2005-2019 IPUMS ACS data.

Figure 2.B7: Greater birthplace contributors to US migrations - temporal mean 2005-2019 share among the US immigrant population

(a) Industry-occupation



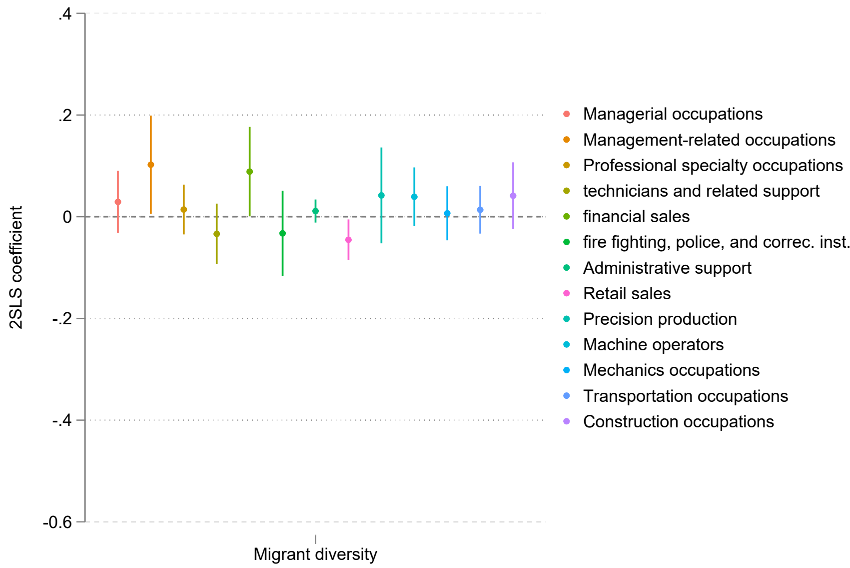
(b) Skill cells



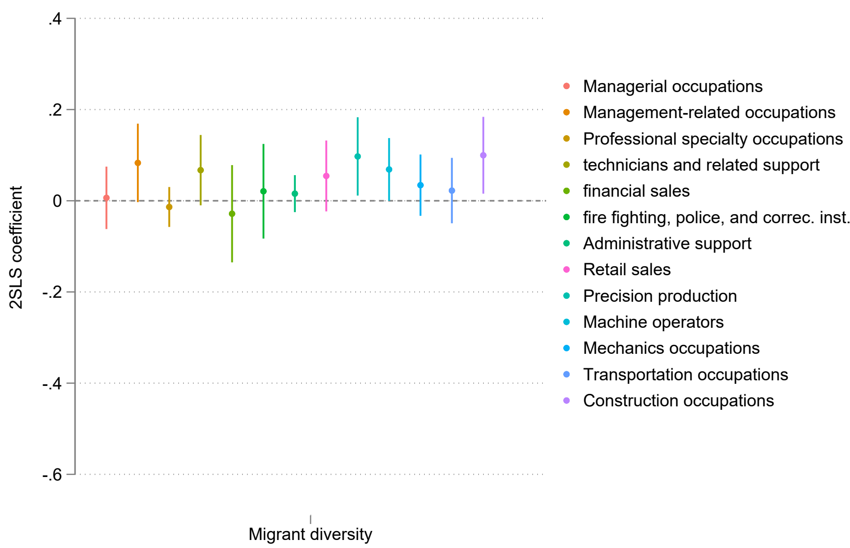
Note: The graphs display 2SLS coefficient and 95% confidence intervals from a regression of the (log of) individual hourly \$1999 wage on a Migrant diversity index computed at metropolitan sector-occupation pairs (a) and at the Ottaviano and Peri (2012) experience-education skill cells (b) levels using aggregated industries samples. It follows eq. (2.3) specification while adding metropolitan total and workers group populations controls. Robust se clustered at metropolitan level. Source: Author's elaboration on IPUMS 2000-2019 ACS data.

Figure 2.B4: IV estimates by industry

(a) Industry-occupation

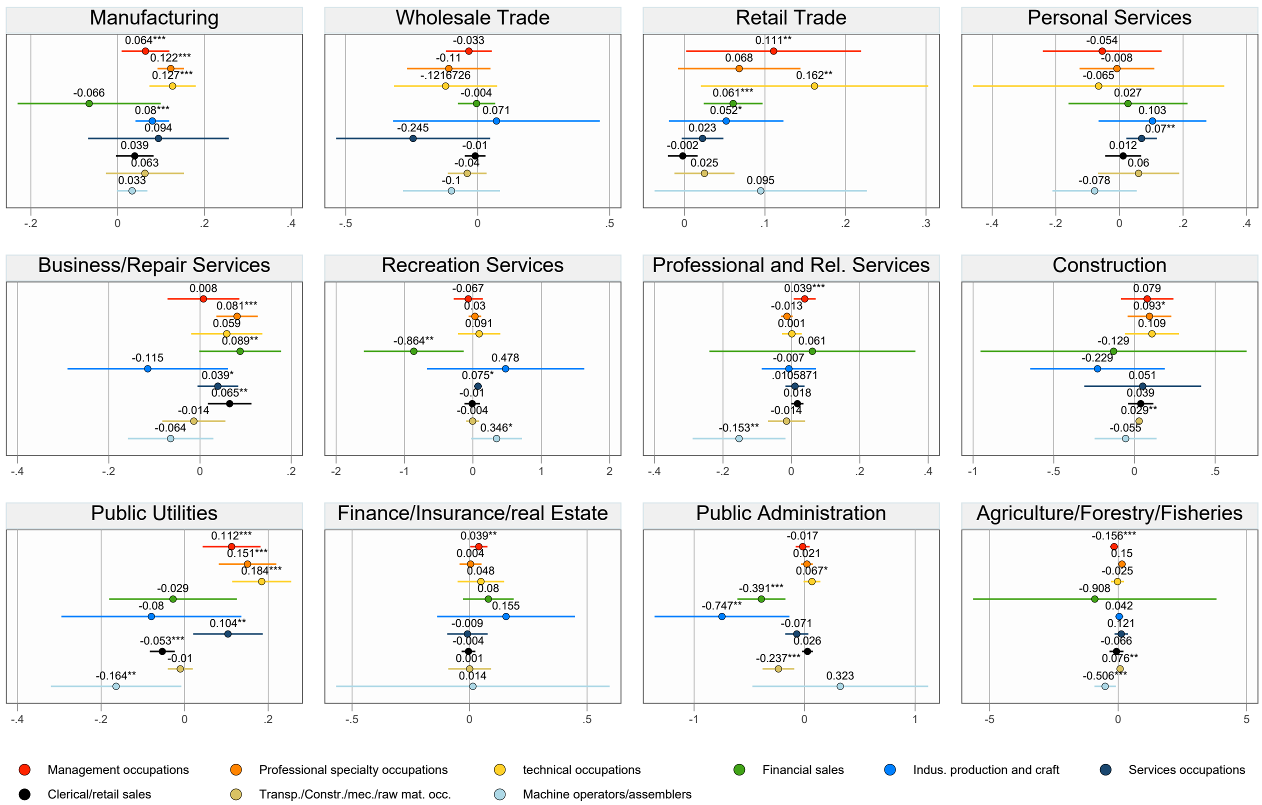


(b) Skill cells



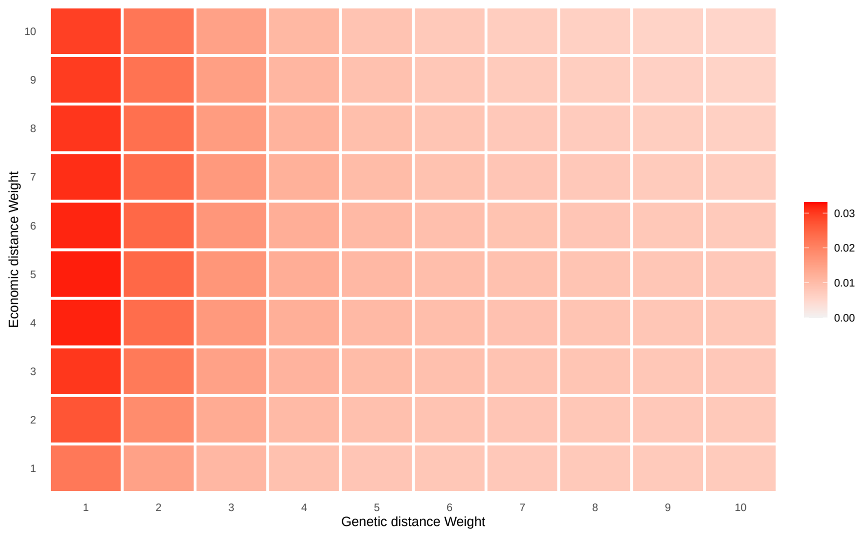
Note: The graphs display 2SLS coefficient and 95% confidence intervals from a regression of the (log of) individual hourly \$1999 wage on a Migrant diversity index computed at metropolitan sector-occupation pairs (a) and at the Ottaviano and Peri (2012) experience-education skill cells (b) levels using aggregated occupations samples. Some aggregated occupations estimates are missing due to their small size or specificity (Extractive occupations, military, etc.). It follows eq. (2.3) specification while adding metropolitan total and workers group populations controls. Robust se clustered at metropolitan level. Source: Author's elaboration on IPUMS 2000-2019 ACS data.

Figure 2.B5: IV estimates by aggregated occupational group



Note: The graphs display OLS coefficient and 95% confidence intervals from a regression of the (log of) individual hourly \$1999 wage on a Migrant diversity index computed at metropolitan sector-occupation pairs by aggregated sector-occupational group samples. It follows eq. (2.3) specification while adding metropolitan total and workers group populations controls. Robust se clustered at the metropolitan level. Source: Author's elaboration on IPUMS 2005-2019 ACS data.

Figure 2.B6: OLS estimates by aggregated sector and occupation



Note: OLS Estimates (eq. (2.3)) with standardised (*z-score*) Augmented Indexes computed at metropolitan sector-occupation pairs level. Economic proximity weights are the absolute difference of the origins 2000-2019 temporal average GDP per capita (2019$), cultural distances weights are Genetic distances from Spolaore and Wacziarg (2016). The reference is the representative US migrant (See fig. 2.B7). Heatplot cells are colored according to the coefficient size. Source: Author's elaboration on IPUMS 2005-2019 ACS, World Bank and Spolaore and Wacziarg (2016) data.

Figure 2.B8: Representative Migrant Augmented Indexes Estimates and Migrant Wages

	B (1)		B (2)		B (3)	
	(1)	(2)	(3)	(4)	(5)	(6)
	migr. div.$_j$	migr. share$_j$	migr. div.$_j$	migr. share$_j$	migr. div$_j$	migr. share$_j$
Pr. mig. div$_j$	0.533***	-0.00665	0.596***	0.00602	0.492***	-0.0130***
	(29.75)	(-1.13)	(28.97)	(0.71)	(29.46)	(-2.76)
Pr. mig. sh.$_j$	-0.105***	0.629***	-0.127***	0.631***	-0.0692*	0.620***
	(-3.87)	(41.65)	(-5.58)	(49.30)	(-1.82)	(37.77)
$ln(Density)$	0.0298*	0.0348**	0.0492**	0.0186	0.0180	0.0458***
	(1.81)	(2.16)	(2.52)	(0.99)	(1.09)	(3.06)
$ln(Density)_j$	0.0740***	0.0105***	0.0635***	0.0113***	0.0812***	0.0104***
	(24.84)	(0.12)	(22.71)	(7.00)	(21.11)	(9.90)
γ_t	Yes	Yes	Yes	Yes	Yes	Yes
γ_c	Yes	Yes	Yes	Yes	Yes	Yes
γ_j	Yes	Yes	Yes	Yes	Yes	Yes
X_{it}	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,049,573	3,049,573	1,137,565	1,137,565	1,909,005	1,909,005

	B (4)		B (5)		B (6)	
	(7)	(8)	(9)	(10)	(11)	(12)
	migr. div$_j$	migr. share$_j$	migr. div$_j$	migr. share$_j$	migr. div$_j$	migr. share$_j$
Pr. migr. div$_j$	0.430***	0.0400***	0.459***	0.0464***	0.361***	0.0356***
	(22.18)	(8.27)	(22.66)	(9.35)	(24.03)	(12.03)
Pr. migr. share$_j$	-0.217***	0.808***	-0.210***	0.823***	-0.0574*	0.800***
	(-7.03)	(38.66)	(-7.48)	(39.66)	(-1.80)	(19.97)
$ln(Density)_j$	0.0912***	-0.0286***	0.0955***	0.00861***	0.100***	0.000841
	(24.51)	(-11.98)	(19.66)	(4.35)	(30.83)	(0.28)
γ_t	Yes	Yes	Yes	Yes	Yes	Yes
γ_c	Yes	Yes	Yes	Yes	Yes	Yes
γ_j	Yes	Yes	Yes	Yes	Yes	Yes
X_{it}	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,408,102	3,408,102	1,288,688	1,288,688	2,119,414	2,119,414

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Instruments are shift-share predicted diversities and migrants shares. Robust se clustered at MSA-level. Source: Author's elaboration on IPUMS 2000-2019 ACS data.

Table 2.B7: First stage regressions of table 2.2 panel B

	Occupational Earnings Quantile			1990 Occupational Earnings Quantile		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Low-skilled	High-skilled	All	Low-skilled	High-skilled
Panel A: OLS estimates						
Migrant diversity$_j$	0.438**	0.177**	0.427**	0.279	0.226	0.294
	(2.17)	(2.33)	(2.01)	(1.49)	(1.18)	(1.42)
Migrant share$_j$	1.531***	0.766	3.447***	1.906***	0.909	4.032***
	(2.20)	(1.61)	(3.93)	(2.93)	(1.85)	(4.88)
$ln(Population)$	-0.495	-0.121	0.218	-0.598	-0.249	0.152
	(-0.67)	(-0.14)	(0.26)	(-0.83)	(-0.27)	(0.18)
$ln(Population_j)$	0.385**	-0.611***	0.447**	0.527***	-0.451***	0.402**
	(2.42)	(-5.66)	(2.55)	(3.86)	(-4.05)	(2.33)
Panel B: IV estimates						
Migrant diversity$_j$	1.362***	1.015**	1.754***	1.007**	0.561	1.540**
	(2.71)	(2.59)	(2.79)	(2.20)	(1.27)	(2.56)
Migrant share$_j$	1.614*	0.0164	4.141***	1.447*	-0.0690	4.601***
	(1.86)	(0.03)	(3.05)	(1.87)	(-0.12)	(3.88)
$ln(Population)$	-0.573	-0.291	0.258	-0.648	-0.385	0.208
	(-0.76)	(-0.33)	(0.29)	(-0.86)	(-0.40)	(0.24)
$ln(Population_j)$	0.325	-0.638***	0.304	0.480***	-0.455***	0.252
	(1.83)	(-5.87)	(1.56)	(3.18)	(-4.07)	(1.34)
Time FE γ_t	Yes	Yes	Yes	Yes	Yes	Yes
Skill Cell FE γ_j	Yes	Yes	Yes	Yes	Yes	Yes
MSA FE γ_c	Yes	Yes	Yes	Yes	Yes	Yes
Individual char. $X_{i,t}$	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,909,823	1,532,502	2,377,321	3,909,823	1,532,502	2,377,321
R-squared	0.319	0.168	0.237	0.294	0.187	0.225
KP F-Stat	249.1	271.1	286.1	249.1	271.1	286.1

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, t statistics in parentheses. Robust se clustered at MSA-level. OLS Panel is a reproduction of table 2.B3 regressions, but using occupational earning quantile defined as the share of the US population having an occupation with a lower median income as the dependent variable. Migrant diversity indexes and shares are computed at metropolitan education-experience skill cell j level. Workers groups are either defined as industry-occupational groups or as Ottaviano and Peri (2012) skill cells. IV panel contains estimates using 2SLS estimates (Migrant share and diversity are instrumented using a shift-share strategy). Dependent variable is log of native \$1999 hourly wage. X_{it} is a vector of individual characteristics containing age, age squared, racial and gender dummies. Source: Author's elaboration on IPUMS 2000-2019 ACS data.

Table 2.B8: OLS and 2SLS estimates by skill groups - workers' education-experience skill cells migrant diversity and occupational earnings

	IV: predicted native migration				IV: predicted external migration			
	Industry-occupation		Skill cells		Industry-occupation		Skill cells	
	Low-skilled	High-skilled	Low-skilled	High-skilled	Low-skilled	High-skilled	Low-skilled	High-skilled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2nd stage estimates (Dep. variable: $ln(wage)$)								
Nat. diversity$_j$	0.00926	0.0275**	-0.00955	-0.121***	-0.0967	0.180**	-0.378	0.117
	(0.88)	(2.12)	(-0.46)	(-4.39)	(-1.00)	(2.05)	(-1.53)	(0.23)
Nat. share$_j$	0.0223**	0.0387***	0.0285	0.102*	0.0242	-0.00251	0.0952	0.0421
	(2.11)	(4.17)	(0.89)	(1.70)	(1.14)	(-0.14)	(1.61)	(0.43)
1st stage estimates (Dep. variable: nat. diversity$_j$)								
pr. migrant div.$_j$	0.483***	0.453***	0.543***	0.443***	0.0598***	-0.0606***	-0.0302***	-0.0124*
	(43.45)	(30.35)	(58.20)	(44.50)	(3.31)	(-2.78)	(-3.58)	(-1.67)
Density variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
γ_t	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
γ_c	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
γ_j	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
X_{it}	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,737,928	2,116,877	1,737,928	2,116,877	1,309,085	2,155,038	1,315,610	2,161,775
K-P F-Stat.	1096.9	645.6	953.1	211.4	10.98	7.708	12.84	2.778

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, t statistics in parentheses. Robust SE clustered at the MSA-level. Intranational Migrant diversity indices and shares are computed at the metropolitan worker group level. Worker groups are either defined as industry-occupational groups or as Ottaviano and Peri (2012) experience-education skill cells. The first four columns report estimates instrumenting native migration variables using (shift-share) predicted internal migrations, while the last four columns report 2SLS estimates instrumenting native migration variables using international predicted migrations. Worker groups are defined either as industry-occupational groups or education-experience skill cells. The dependent variable is the log of the native individual hourly wage. Density variables include metropolitan and metropolitan-group densities. X_{it} includes age, age squared, race, and gender controls. Source: Author's elaboration on IPUMS 2000–2019 ACS data.

Table 2.B9: 2SLS estimates - US internal migration diversity and wages

	Industry-occupation				Education-Experience Skill cells			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Low-skilled	High-skilled	Low-skilled	High-skilled	Low-skilled	High-skilled	Low-skilled	High-skilled
Migrant diversity$_j$			0.0154*	0.00734			0.0224	-0.0130
			(1.76)	(0.72)			(1.28)	(-0.68)
Migrant share$_j$			-0.148***	-0.125***			-0.0764***	0.0614
			(-7.53)	(-3.24)			(-3.15)	(1.00)
Nat. diversity$_j$	-0.0967	0.180**	0.00158	0.0234***	-0.378	0.117	-0.00716	-0.0587***
	(-1.00)	(2.05)	(0.30)	(3.13)	(-1.53)	(0.23)	(-0.61)	(-4.38)
Nat. share$_j$	0.0242	-0.00251	0.00487	0.0317***	0.0952	0.0421	0.0137	0.0753***
	(1.14)	(-0.14)	(1.16)	(6.50)	(1.61)	(0.43)	(0.68)	(3.55)
1st stage estimates								
pr. migrant div.$_j$	0.0598***	-0.0606***	0.610***	0.534***	-0.0302***	-0.0124*	0.425***	0.356***
	(3.31)	(-2.78)	(32.92)	(37.97)	(-3.58)	(-1.67)	(27.91)	(26.84)
Nat. migration instrumented	Yes	Yes	No	No	Yes	Yes	No	No
Density variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
γ_t	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
γ_c	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
γ_j	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
X_{it}	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,309,085	2,155,038	1,137,565	1,909,005	1,315,610	2,161,775	1,288,688	2,119,414
K-P F-Stat.	10.98	7.708	386.8	395.3	12.84	2.778	419.4	126.7

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, t statistics in parentheses. Robust SE clustered at the MSA-level. An intranational immigrant is defined here as a worker born in a U.S. state other than the U.S. state in which he lives. Intranational Migrant diversity indices and shares are computed at the metropolitan worker group level. Worker groups are either defined as industry-occupational groups or as Ottaviano and Peri (2012) experience-education skill cells. Columns displaying Migrant diversity estimates instruments Migrant share and diversity using a shift-share strategy, while native migration variables are not instrumented and serve as controls. Other columns use a shift-share approach to instrument native diversity and share. The dependent variable is the log of the native $1999 individual hourly wage. Density variables include the (log of) metropolitan and metropolitan group j densities. X_{it} is a vector of individual characteristics containing age, age squared, racial, and gender dummies. Source: Author's elaboration on IPUMS 2000-2019 ACS data.

Table 2.B10: 2SLS estimates by skill group: intranational migration and wages

	abstract/routine			abstract/manual		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Low-skilled	High-skilled	All	Low-skilled	High-skilled
Panel A: OLS estimates						
migrant div.$_j$	-0.0105	0.00256	-0.0469***	2.135	-0.537	4.594*
	(-1.10)	(0.30)	(-3.65)	(1.17)	(-0.69)	(1.83)
Migrant share$_j$	0.0760**	0.0211	0.245***	-9.089**	1.802	-5.522
	(2.32)	(0.58)	(5.28)	(-2.39)	(1.07)	(-0.95)
$ln(density)$	-0.0355	-0.0508**	-0.0238	-5.936**	-2.139	-5.606
	(-1.53)	(-2.15)	(-0.74)	(-2.12)	(-0.67)	(-1.46)
$ln(density_j)$	0.00398	-0.0343***	0.00266	9.496***	-0.297	14.34***
	(0.41)	(-3.61)	(0.15)	(3.12)	(-0.42)	(3.17)
Panel B: IV estimates						
migrant div.$_j$	0.0312	0.0693***	-0.0315	1.774	-3.345*	5.932
	(1.54)	(2.97)	(-0.83)	(0.44)	(-1.81)	(0.85)
Migrant share$_j$	0.0703**	0.0219	0.235***	-9.041**	1.770	-6.328
	(2.09)	(0.57)	(4.41)	(-2.43)	(1.06)	(-0.98)
$ln(density)$	-0.0393*	-0.0625**	-0.0235	-5.903**	-1.645	-5.576
	(-1.68)	(-2.50)	(-0.73)	(-2.07)	(-0.51)	(-1.46)
$ln(density_j)$	0.00157	-0.0368***	0.00136	9.516***	-0.191	14.23***
	(0.16)	(-4.03)	(0.08)	(3.08)	(-0.28)	(3.10)
Time FE γ_t	Yes	Yes	Yes	Yes	Yes	Yes
Work. groups FE γ_j	Yes	Yes	Yes	Yes	Yes	Yes
MSA FE γ_c	Yes	Yes	Yes	Yes	Yes	Yes
Individual char. $X_{i,t}$	Yes	Yes	Yes	Yes	Yes	Yes
Dep. var. SD	1.30	1.05	1.40	185.42	139.82	218.55
Observations	3,989,494	1,564,691	2,424,803	3,945,763	1,555,254	2,390,509
KP F-Stat	488.7	669.5	432.4	489.5	668.6	434.7

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, t statistics in parentheses. Robust SE clustered at the MSA-level. Panel A displays OLS estimates based on eq. (2.3) by skill groups. Migrant diversity indices and shares are computed at the metropolitan worker experience-education skill cell level. Panel B contains 2SLS estimates (Migrant share and diversity are instrumented using a shift-share strategy). The dependent variable is the ratio of occupational task content. X_{it} is a vector of individual characteristics containing age, age squared, racial, and gender dummies. Source: Author's elaboration on IPUMS 2000-2019 ACS data.

Table 2.B11: OLS and 2SLS estimates - experience-education cells immigrant diversity and native occupational tasks

2.B.4 Workers groups definition

All 1990-defined occupations and industries are listed here. In this article, a workers group is defined as workers belonging to the same 1990 industry and the same 1990 occupation. Occupations are sorted here according to David and Dorn (2013) aggregated groups definitions and industries according to IPUMS aggregated industries.

Occupations

Executive, Administrative, and Managerial Occupations: Legislators, Chief executives and public administrators, Financial managers, Human resources and labor relations managers, Managers and specialists in marketing, advertising, and public relations, Managers in education and related fields, Managers in education and related fields, Managers of medicine and health occupations, Postmasters and mail superintendents, Managers of food-serving and lodging establishments, Managers of properties and real estate, Funeral directors, Managers of service organisations, n.e.c., Managers and administrators, n.e.c.

Management Related Occupations: Accountants and auditors, Insurance underwriters, Other financial specialists, Management analysts, Personnel, HR, training, and labor relations specialists, Purchasing agents and buyers of farm products, Buyers, wholesale and retail trade, Purchasing managers, agents and buyers, n.e.c., Business and promotion agents, Construction inspectors, Inspectors and compliance officers (outside construction), Management support occupations.

Professional specialty occupations: Architects, Engineers (Aerospace engineer, Metallurgical and materials engineers (variously phrased), Petroleum, mining, and geological engineers, Chemical engineer, Civil engineers, Electrical engineer, Industrial engineers, Mechanical engineers, Not-elsewhere-classified engineers), Mathematical and Computer Scientists (Computer systems analysts and computer scientists, Operations and systems researchers and analysts, Actuaries, Statisticians, Mathematicians and mathematical scientists), Natural Scientists (Physicists and astronomers, Chemists, Atmospheric and space scientists, Geologists, Physical scientists n.e.c., Agricultural and food scientists, Biological scientists, Foresters and conservation scientists, Medical scientists), Health Diagnosing Occupations (Physicians, Dentists, Veterinarians, Optometrists, Podiatrists, Other health and therapy), Health Assessment and Treating Occupations (Registered nurses, Pharmacists, Dietitians and nutritionists), Therapists (Respiratory therapists, Occupational therapists, Physical therapists, Speech therapists, Therapists n.e.c., Physicians' assistants), Postsecondary Teachers (Subject instructors (HS/college)), Teachers Except Postsecondary (Kindergarten and earlier school teachers, Primary school teachers, Secondary school teachers, Special education teachers, Teachers n.e.c, Vocational and educational counselors), Librarians, Archivists and Curators, Social Scientists and Urban Planners (Economists, market researchers, and survey researchers, Psychologists, Sociologists, Social scientists n.e.c., Urban and regional planners), Lawyers and Judges, Writers, Artists, Entertainers, and Athletes (Writers and authors, Technical writers, Designers, Musician or composer, Actors, directors, producers, Art makers (painters, sculptors, craft-artists, and print-makers), Photographers, Dancers, Art/entertainment performers and related, Editors and reporters, Announcers, Athletes, sports instructors, and officials).

Technicians and Related Support Occupations: Health Technologists and Technicians (Clinical laboratory technologies and technicians, Dental hygienists, Health record tech specialists, Radiologic tech specialists, Licensed practical nurses, Health technologists and technicians n.e.c.), Engineering and Related Technologists and Technicians (Electrical and electronic (engineering) technicians, Engineering technicians n.e.c., Mechanical engineering technicians, Drafters, Surveyors, cartographers, mapping scientists and technicians, Biological technicians), Science Technicians (Chemical technicians, Other science technicians), Technicians, Except Health, Engineering, and Science (Airplane pilots and navigators, Air traffic controllers, Broadcast equipment operators, Computer software developers, Programmers of numerically controlled machine tools, Legal assistants, paralegals, legal support, etc., Technicians, n.e.c.).

financial sales and related Occupations: Supervisors and proprietors of sales jobs, Sales Representatives, Finance and Business Services (Insurance sales occupations, Real estate sales occupations, Financial services sales occupations, Advertising and related sales jobs), Sales engineers.

fire fighting, police, and correctional institutions Occupations: Fire fighting, prevention, and inspection police and detectives, other Police and Detectives (Police, detectives, and private investigators, Other law enforcement: sheriffs, bailiffs, correctional institution officers).

Low-skilled occupations

administrative support and retail sales occupations: Salespersons, n.e.c., Retail sales clerks, Cashiers, Door-to-door sales, street sales, and news vendors, Sales demonstrators/promoters/models, Administrative Support Occupations, Including Clerical, Office supervisors, Computer and peripheral equipment operators, Secretaries, Stenographers, Typists, Interviewers, enumerators, and surveyors, Hotel clerks, Transportation ticket and reservation agents, Receptionists, Information clerks, n.e.c., Correspondence and order clerks, Human resources clerks, except payroll and timekeeping, Library assistants, File clerks, Records clerks, Bookkeepers and accounting and auditing clerks, Payroll and timekeeping clerks, Cost and rate clerks (financial records processing), Billing clerks and related financial records processing, Duplication machine operators/office machine operators, Mail and paper handlers, Office machine operators, n.e.c., Telephone operators, Other telecom operators, Postal clerks, excluding mail carriers, Mail carriers for postal service, Mail clerks, outside of post office, Messengers, Dispatchers, Inspectors, n.e.c., Shipping and receiving clerks, Stock and inventory clerks, Meter readers, Weighers, measurers, and checkers, Material recording, scheduling, production, planning, and expediting clerks, Insurance adjusters, examiners, and investigators, Customer service reps, investigators and adjusters, except insurance, Eligibility clerks for government programs; social welfare, Bill and account collectors, General office clerks, Bank tellers, Proofreaders, Data entry keyers, Statistical clerks, Teacher's aides, Administrative support jobs, n.e.c.

Low-skill services: Housekeepers, maids, butlers, stewards, and lodging quarters cleaners, Private household cleaners and servants, Protective Service Occupations, Supervisors of guards, Crossing guards and bridge tenders, Guards, watchmen, doorkeepers, Protective services, n.e.c., Bartenders, Waiter/waitress, Cooks, variously defined, Food counter and fountain workers, Kitchen workers, Waiter's assistant, Misc. food prep workers, Dental assistants, Health aides, except nursing, Nursing aides, orderlies, and attendants, Supervisors of cleaning and building service, Janitors, Elevator operators, Pest control occupations, Barbers, Hairdressers and cosmetologists, Recreation facility attendants, Guides, Ushers, Public transportation attendants and inspectors, Baggage porters, Welfare service aides, Child care workers, Personal service occupations, n.e.c.

Precision production and craft occupations: Tool and die makers and die setters, Machinists, Boilermakers,Precision grinders and filers, Patternmakers and model maker, Lay-out workers, Engravers, Tinsmiths, coppersmiths, and sheet metal workers, Cabinetmakers and bench carpenters, Furniture and wood finishers, Other precision woodworkers, Dressmakers and seamstresses, Tailors, Upholsterers, Shoe repairers, Other precision apparel and fabric workers, Hand molders and shapers, except jewelers, ptical goods workers, Dental laboratory and medical appliance technicians, Bookbinders, Other precision and craft workers, Butchers and meat cutters, Bakers, Batch food makers, Adjusters and calibrators, Water and sewage treatment plant operators, Power plant operators, Plant and system operators, stationary engineers, Other plant and system operators.

Machine operators, assemblers and inspectors: Lathe, milling, and turning machine operatives, Punching and stamping press operatives, Rollers, roll hands, and finishers of metal, Drilling and boring machine operators, Grinding, abrading, buffing, and polishing workers, Forge and hammer operators, Fabricating machine operators, n.e.c., Molders, and casting machine operators, Metal platers, Heat treating equipment operators, Wood lathe, routing, and planing machine operators, Sawing machine operators and sawyers, Shaping and joining machine operator (woodworking), Nail and tacking machine operators (woodworking), Other woodworking machine operators, Printing machine operators, n.e.c., Photoengravers and lithographer, Winding and twisting textile/apparel, Knitters, loopers, and toppers textile operatives, Textile cutting machine operators, Textile sewing machine operators, Shoemaking machine operators, Pressing machine operators (clothing), Laundry workers, Misc. textile machine operators, Cementing and gluing machine operators, Packers, fillers, and wrappers, Extruding and forming machine operators, Mixing and blending machine operatives, Separating, filtering, and clarifying machine operators, Painting machine operators, Roasting and baking machine operators (food), Washing, cleaning, and pickling machine operators, Paper folding machine operators, Furnace, kiln, and oven operators, apart from food, Crushing and grinding machine operator, Slicing and cutting machine operators, Motion picture projectionists, Photographic process workers, Machine operators, n.e.c., Welders and metal cutters, Solderers, Assemblers of electrical equipment, Hand painting, coating, and decorating occupations, Production checkers and inspectors,Graders and sorters in manufacturing.

transportation/construction/mechanics/mining/agricultural occupations: Farmers (owners and tenants), Horticultural specialty farmers, Farm managers, except for horticultural farms, Farm workers, Marine life cultivation workers, Nursery farming workers, Supervisors of agricultural occupations, Gardeners and groundskeepers, Animal caretakers except on farms, Graders and sorters of agricultural products, Inspectors of agricultural products, Timber, logging, and forestry workers, Fishers, hunters, and kindred Supervisors of mechanics and repairers, Mechanics and Repairers, Except Supervisors, Automobile mechanics, Bus, truck, and stationary engine mechanic, Aircraft mechanics, Small engine repairers, Auto body repairers, Heavy equipment and farm equipment mechanics, Industrial machinery repairers, Machinery maintenance occupations, Repairers of industrial electrical equipment, Repairers of data processing equipment, Repairers of household appliances and power tools, Telecom and line installers and repairers, Repairers of electrical equipment, n.e.c., Heating, air conditioning, and refrigeration mechanics, Precision makers, repairers, and smiths,Locksmiths and safe repairers, Office machine repairers and mechanics, Repairers of mechanical controls and valves, Elevator installers and repairers, Millwrights, Mechanics and repairers, n.e.c., Supervisors of construction work, Masons, tilers, and carpet installers, Carpenters, Drywall installers, Electricians, Electric power installers and repairers, Painters, construction and maintenance, Paperhangers, Plasterers, Plumbers, pipe fitters, and steamfitters, Concrete and cement workers, Glaziers, Insulation workers, Paving, surfacing, and tamping equipment operators, Roofers and slaters, Sheet metal duct installers, Structural metal workers, Drillers of earth, Construction trades, n.e.c., Drillers of oil wells, Explosives workers, Miners, Other mining occupations, Supervisors of motor vehicle transportation, Truck, delivery, and tractor drivers, Bus drivers, Taxi cab drivers and chauffeurs, Parking lot attendants, Railroad conductors and yardmasters, Locomotive operators (engineers and firemen), Railroad brake, coupler, and switch operators, Ship crews and marine engineers, Water transport infrastructure tenders and crossing guards, Operating engineers of construction equipment, Crane, derrick, winch, and hoist operators,Excavating and loading machine operators, Misc. material moving occupations, Helpers, constructions, Helpers, surveyors, Construction laborers, Production helpers, Garbage and recyclable material collectors, Materials movers: stevedores and longshore workers, Stock handlers, Machine feeders and offbearers, Freight, stock, and materials handlers, Garage and service station related occupations, Vehicle washers and equipment cleaners, Packers and packagers by hand, Laborers outside construction.

Industries

Agriculture, Forestry, fisheries and Mining: Agricultural production: crops and livestock, Veterinary services, Landscape and horticultural services, Agricultural services, n.e.c., Forestry, Fishing, hunting, and trapping, Metal mining, Coal mining, Oil and gas

Occupational group	hourly wage	Prestige score	Earning quant.	Status Score
Farm operators and managers	8.99	41.40	37.42%	40.40
Other agricultural and related occupations	9.11	27.70	20.02%	13.55
Low-skilled services	10.23	14.46	17.42%	20.11
Transportation and material moving	12.75	29.98	32.06%	26.63
machine operators, assemblers, and inspectors	13.78	33.31	33.99%	27.96
Retail sales	14.12	32.70	27.28%	34.17
Administrative support	14.47	39.84	34.08%	43.70
Construction	14.75	41.86	44.82%	34.31
Precision production	15.76	44.74	47.33%	43.49
Mechanics and repairers	16.16	40.61	52.48%	45.63
Extractive occupations	17.59	34.13	66.69%	43.73
Financial sales and related occupations	20.59	45.33	55.74%	62.83
Fire fighting, police, and correctional institutions	22.81	54.98	76.34%	70.76
Technicians and related support	23.94	56.53	68.32%	72.43
Professional specialty occupations	24.05	62.89	64.40%	76.73
management-related occupations	26.35	54.62	74.35%	80.70
Executive, administrative and managerial occ.	30.27	56.36	84.10%	82.23

Note: wage is the average metropolitan individual wage in the occupational group. Prestige is the (population-weighted) Nakao and Treas (1994) average prestige score of the occupation. All the means are computed using the US metropolitan population. Earning quantile represents the share of US population that has a lower total annual income lower than the average (population-weighted) occupation in the occupational group. The Nam and Boyd (2004) status score is a composite measure related to both the median earnings and median educational attainment associated to an occupation.

Table 2.B12: Occupational groups characteristics

extraction, Nonmetallic mining and quarrying, except fuels.

Construction: All Construction.

Manufacturing: Meat products, Dairy products, Canned, frozen, and preserved fruits and vegetables, Grain mill products, Bakery products, Sugar and confectionery products, Beverage industries, Misc. food preparations and kindred products, Food industries, n.s., Tobacco manufactures, Knitting mills, Dyeing and finishing textiles, except wool and knit goods, Carpets and rugs, Yarn, thread, and fabric mills, Miscellaneous textile mill products, Apparel and accessories, except knit, Miscellaneous fabricated textile products, Pulp, paper, and paperboard mills, Miscellaneous paper and pulp products, Paperboard containers and boxes, Printing, publishing, and allied industries, Plastics, synthetics, and resins, Drugs, Soaps and cosmetics, Paints, varnishes, and related products, Agricultural chemicals, Industrial and miscellaneous chemicals, Petroleum refining, Miscellaneous petroleum and coal products, Tires and inner tubes, Other rubber products, and plastics footwear and belting, Miscellaneous plastics products, Leather tanning and finishing, Footwear, Leather products, Lumber and woods products, Logging, Sawmills, planing mills, and millwork, Wood buildings and mobile homes, Miscellaneous wood products, Furniture and fixtures, Glass and glass products, Cement, concrete, gypsum, and plaster products, Structural clay products, Pottery and related products, Misc. nonmetallic mineral and stone products, Blast furnaces, steelworks, rolling and finishing mills, Iron and steel foundries, Primary aluminum industries, Other primary metal industries, Cutlery, handtools, and general hardware, Fabricated structural metal products, Screw machine products, Metal forgings and stampings, Ordnance, Miscellaneous fabricated metal products, Metal industries, Engines and turbines, Farm machinery and equipment, Construction and material handling machines, Metalworking machinery, Office and accounting machines, Computers and related equipment, Machinery, except electrical, n.e.c., Machinery, n.s., Household appliances, Radio, TV, and communication equipment, Electrical machinery, equipment, and supplies, n.e.c., Electrical machinery, equipment, and supplies, n.s., Motor vehicles and motor vehicle equipment, Aircraft and parts, Ship and boat building and repairing, Railroad locomotives and equipment, Guided missiles, space vehicles, and parts, Cycles and miscellaneous transportation equipment, Scientific and controlling instruments, Medical, dental, and optical instruments and supplies, Photographic equipment and supplies, Watches, clocks, and clockwork operated devices, Toys, amusement, and sporting goods, Miscellaneous manufacturing industries, Manufacturing industries, n.s.

Technicians and Related Support Occupations: Health Technologists and Technicians (Clinical laboratory technologies and technicians, Dental hygienists, Health record tech specialists, Radiologic tech specialists, Licensed practical nurses, Health technologists and technicians n.e.c.), Engineering and Related Technologists and Technicians (Electrical and electronic (engineering) technicians, Engineering technicians n.e.c., Mechanical engineering technicians, Drafters, Surveyors, cartographers, mapping scientists and technicians, Biological technicians), Science Technicians (Chemical technicians, Other science technicians), Technicians, Except Health, Engineering, and Science (Airplane pilots and navigators, Air traffic controllers, Broadcast equipment operators, Computer software developers, Programmers of numerically controlled machine tools, Legal assistants, paralegals, legal support, etc., Technicians, n.e.c.).

Public Utilities: Railroads, Bus service and urban transit, Taxicab service, Trucking service, Warehousing and storage, U.S. Postal Service, Water transportation, Air transportation, Pipe lines, except natural gas, Services incidental to transportation, Radio and television broadcasting and cable, Telephone communications, Telegraph and miscellaneous communications services, Electric light and power, Gas and steam supply systems, Electric and gas, and other combinations, Water supply and irrigation, Sanitary services, Utilities, n.s.

Wholesale Trade: Motor vehicles and equipment, Furniture and home furnishings, Lumber and construction materials, Professional and commercial equipment and supplies, Metals and minerals, except petroleum, Electrical goods, Hardware, plumbing and heating supplies, Machinery, equipment, and supplies, Scrap and waste materials, Miscellaneous wholesale - durable goods, Paper and paper products, Drugs, chemicals, and allied products, Apparel, fabrics, and notions, Groceries and related products, Farm-product raw materials, Petroleum products, Alcoholic beverages, Farm supplies, Miscellaneous wholesale - nondurable goods, Wholesale trade, n.s.

Retail trade: Lumber and building material retailing, Hardware stores, Retail nurseries and garden stores, Mobile home dealers, Department stores, Variety stores, Miscellaneous general merchandise stores, Grocery stores, Dairy products stores, Retail bakeries, Food stores, n.e.c., Motor vehicle dealers, Auto and home supply stores, Gasoline service stations, Miscellaneous vehicle dealers, Apparel and accessory stores, Shoe store,Furniture and home furnishings stores, Household appliance stores, Radio, TV, and computer stores, Music stores, Eating and drinking places, Drug stores, Liquor stores, Sporting goods, bicycles, and hobby stores, Book and stationery stores, Jewelry stores, Gift, novelty, and souvenir shops, Sewing, needlework, and piece goods stores, Catalog and mail order houses, Vending machine operators, Direct selling establishments, Fuel dealers, Retail florists, Miscellaneous retail stores, Retail trade, n.s.

Finance, Insurance and Real estate: Banking, Savings institutions, including credit unions, Credit agencies, n.e.c., Security, commodity brokerage, and investment companies, Insurance, Real estate, including real estate-insurance offices.

Business and Repair services: Advertising, Services to dwellings and other buildings, Personnel supply services, Computer and data processing services, Detective and protective services, Business services, n.e.c., Automotive rental and leasing, without drivers, Automobile parking and carwashes, Automotive repair and related services, Electrical repair shops, Miscellaneous repair services.

Personal Services: Private households, Hotels and motels, Lodging places, Laundry, cleaning, and garment services, Beauty shops, Barber shops, Funeral service and crematories, Shoe repair shops, Dressmaking shops, Miscellaneous personal services.

Recreation Services: Theaters and motion pictures, Video tape rental, Bowling centers, Miscellaneous entertainment and recreation services.

Professional and Related Services: Offices and clinics of physicians, Offices and clinics of dentists, Offices and clinics of chiropractors, Offices and clinics of optometrists, Offices and clinics of health practitioners, n.e.c., Hospitals, Nursing and personal care facilities, Health services, n.e.c., Legal services, Elementary and secondary schools, Colleges and universities, Vocational schools, Libraries, Educational services, n.e.c., Job training and vocational rehabilitation services, Child day care services, Family child care homes, Residential care facilities, without nursing, Social services, n.e.c., Museums, art galleries, and zoos, Labor unions, Religious organizations, Membership organizations, n.e.c., Engineering, architectural, and surveying services, Accounting, auditing, and bookkeeping services, Research, development, and testing services, Management and public relations services, Miscellaneous professional and related services.

Public Administration: Executive and legislative offices, General government, n.e.c., Justice, public order, and safety, Public finance, taxation, and monetary policy, Administration of human resources programs, Administration of environmental quality and housing programs, Administration of economic programs, National security and international affairs.

Military: Army, Air Force, Navy, Marines, Coast Guard, Armed Forces, branch not specified, Military Reserves or National Guard.

2.B.5 An illustrative model of amenities heterogeneous preferences

This section extends the seminal model of Ottaviano and Peri (2006) by adding worker heterogeneity in terms of skills and preferences in accordance with theories from other disciplines (Clark et al., 2002; Florida et al., 2008). It builds on the assumptions of perfect geographic mobility, occupational immobility, and heterogeneous preferences. There are K types of workers indexed k with their specific preferences with respect to local amenities, living in a world composed of C regions indexed c. I consider the representative household for each class of workers k in the area c whose preferences can be represented by the following utility function:

$$U_{c,k} = A_c^{U_k}(a_c) Y_{c,k}^{\mu} H_{c,k}^{1-\mu} \tag{2.B1}$$

Where $Y_{c,k}$ represents the consumption of the aggregated good and $H_{c,k}$ the consumption of land. $A_c^{U_k}(a_c)$ is a worker-class-specific function capturing the taste for local amenities (vector a_c) including a local diversity of people. Note that I do not make any assumption yet on the sign of the differential with respect to diversity for any class of workers. The Utility Maximisation Problem results in the following classic demand functions:

$$\begin{cases} Y_{c,k} = \frac{\mu I_{c,k}}{P_c} \\ H_{c,k} = \frac{(1-\mu) I_{c,k}}{r_c} \end{cases} \tag{2.B2}$$

Where $I_{c,k}$ is the household income while P_c and r_c are respectively the aggregated good price and the rent from land in region c. The indirect utility function reads $V_{c,k}(I_{c,k}, P_c, r_c, a_c) = A_c^{U_k}(a_c)\mu^{\mu}(1-\mu)^{1-\mu} P_c^{-\mu} r_c^{\mu-1} I_{c,k}$. Considering $B = \mu^{\mu}(1-\mu)^{1-\mu}$, I obtain:

$$V_{c,k}(I_{c,k}, P_c, r_c, a_c) = B A_c^{U_k}(a_c) P_c^{-\mu} r_c^{\mu-1} I_{c,k} \tag{2.B3}$$

I assume free mobility across regions, implying that the representative household freely chooses the location c, maximizing eq. (2.B1). As a consequence, a spatial equilibrium requires the value of the indirect utility to be equal for all geographic areas c. Turning to the production part, I consider homogeneous firms producing the aggregated good using land and labor under perfect competition; I consider that each area has a fixed supply of land $H_c = \sum_i n_i H_{ci} + \sum_j H_{cj}$ where i indexes workers' categories and j indexes the firms. I consider $n_i = 1 \ \forall i$. The Cobb-Douglas production function for a firm j producing in area c reads:

$$Y_{cj} = A_c^Y(a_c) H_{cj}^{1-\sum_i \alpha_i} \prod_i L_{cji}^{\alpha_i}, \quad with\ \alpha_i \in]0,1[\ \forall i, \quad \sum_i \alpha_i \leq 1 \tag{2.B4}$$

Where $L_{c,j,i}$ is the amount of labour from class of worker i used by firm j in the area c, $w_{c,i}$ is the local wage of category i and $H_{c,j}$ is the amount of land used by the firm j in c in order to produce. The Costs Minimisation Problem leads to the following k-type labour conditional demand:

$$L_{c,j,k} = \frac{\alpha_k}{1-\sum_i \alpha_i} \frac{r_c}{w_{c,k}} H_{c,j} \tag{2.B5}$$

Plugging this in the cost function and defining $W_c = r_c^{1-\sum_{i=1}^K \alpha_i} \prod_{i=1}^K w_{c,i}^{\alpha_i}$ and $\Lambda = (1-\sum_{i=1}^K \alpha_i)^{1-\sum_{i=1}^K \alpha_i} \prod_{i=1}^K \alpha_i^{\alpha_i}$, one obtains the following cost function associated to a defined level of production $\bar{Y_{cj}}$:

$$C_{c,j}(W_c, \bar{Y_{cj}}) = A_c^Y(a_c)^{-1} \frac{W_c}{\Lambda} \bar{Y_{cj}} \tag{2.B6}$$

Plugging the eq. (2.B6) total cost expression in the profit maximisation problem of a firm, the first order condition allows to express the aggregated good price as a function of regional factors, incomes, and amenities:

$$P_c \equiv \frac{W_c}{A_c^Y(a_c)\Lambda} = \frac{r_c^{1-\sum \alpha_i} \prod_i^I w_{ci}^{\alpha_i}}{A_c^Y(a_c)(1-\sum \alpha_i)^{1-\sum \alpha_i} \prod_i^K \alpha_i^{\alpha_i}} \tag{2.B7}$$

To go further, I follow Ottaviano and Peri (2006) by making the assumption that the aggregated good is freely tradable across areas, implying the same price everywhere that I normalize at $P_c \equiv P = 1$. From the previous eq. (2.B7), the free trade condition therefore implies $A_c^Y(a_c)\Lambda = W_c$. This is the spatial equilibrium condition related to trade; Another spatial equilibrium condition related to amenities comes from the eq. (2.B3) indirect utility equality. With regard to this condition, I firstly consider regional land and labor market clearing conditions:

$$\begin{cases} L_{ck} = \sum_j L_{cjk} \\ H_c = \sum_j H_{cj} + \sum_i H_{ci} \end{cases} \tag{2.B8}$$

Plugging conditional demands (eq. (2.B5)) inside, I obtain:

$$\begin{cases} L_{ck} = \sum_j^J \frac{\alpha_k}{\Lambda} \frac{W_c}{w_{ck}} \frac{Y_{cj}}{A_c(a_c)} \\ H_c = \sum_j^J \frac{(1-\sum_i \alpha_i)}{\Lambda} \frac{W_c}{r_c} \frac{Y_{cj}}{A_c(a_c)} + \sum_i^N H_{ci} \end{cases} \tag{2.B9}$$

Using demands from eq. (2.B2) and using the condition from eq. (2.B7), one can find the labor and total housing demands in region c:

$$\begin{cases} L_{c,k} = \alpha_k \frac{P_c Y_c}{w_{c,k}}, \forall k \in \{1, ..., K\} \\ H_c = (1 - \sum_i \alpha_i + \frac{1-\mu}{\mu}) \frac{P_c Y_c}{r_c} \end{cases} \tag{2.B10}$$

Before going any further, I assume that the representative worker's endowment consists in one unit of labor that he allocates entirely to the production of the aggregate good: eq. (2.B3) becomes $V_k = BA_c^{U_k}(a_c) r_c^{\mu-1} w_{c,k}$. I plug eq. (2.B10) rent and wage expressions in the indirect utility function, considering the constant $D = \mu^\mu (\frac{(1-\mu)\mu}{(1-\mu)+\mu(1-\sum_i \alpha_i)})^{1-\mu}$. By doing so, I obtain a new expression of the indirect utility function, $V_k = DA_c^{U_k}(a_c) Y_c^\mu H_c^{1-\mu} L_{c,k}^{-1}$, that allows us to retrieve the following conditional demand for labor k under free trade and free migration assumptions:

$$L_{c,k} = DA_c^{U_k}(a_c) \alpha_k Y_c^\mu V_k^{-1} \tag{2.B11}$$

I consider the ratio of such k-labor demands of 2 regions c and c':

$$\frac{L_{c,k}}{L_{c',k}} = \frac{A_c^{U_k}(a_c) Y_c^\mu H_c^{1-\mu}}{A_{c'}^{U_k}(a_{c'}) Y_{c'}^\mu H_{c'}^{1-\mu}} \tag{2.B12}$$

By multiplying the labor ratios of the different types of labor together, with the appropriate exponents:

$$\prod_i \left(\frac{L_{c,i}}{L_{c',i}}\right)^{\alpha_i} = \prod_i \left(\frac{A^{U_k}_{c,i}(a_c)}{A^{U_k}_{c',i}(a_{c'})}\right)^{\alpha_i} \left(\frac{Y_c}{Y_{c',i}}\right)^{\mu\alpha_i} \left(\frac{H_c}{H_{c'}}\right)^{(1-\mu)\alpha_i}$$

After replacing regional labor products using eq. (2.B4) and considering the homogeneity of firms, one obtains after some computation[24]:

$$\left(\frac{Y_c}{Y_{c'}}\right)^{1-\mu\sum_i \alpha_i} = \left(\frac{H_c}{H_{c'}}\right)^{1-\mu\sum_i \alpha_i} \frac{A^Y_c(a_c)}{A^Y_{c'}(a_{c'})} \prod_i \left(\frac{A^{U_i}_{c,i}(a_c)}{A^{U_i}_{c',i}(a_{c'})}\right)^{\alpha_i}$$

One can plug such expression in eq. (2.B12) as well as the national employment of type k $L_k = \sum_c L_{c,k}$, considering $\kappa_k = \sum_c A^{U_k}_{c,k}(a_c) H_c (A^Y_{c'})^{\frac{\mu}{1-\mu\sum_i \alpha_i}} (\prod_i A^{U_i}_{c,i}(a_c))^{\frac{\mu}{1-\mu\sum_i \alpha_i}}$ in order to retrieve the region c' employment of type k:

$$L_{c',k} = A^{U_k}_{c,k}(a_c) H_{c'} A^Y_{c'}(a_c)^{\frac{\mu}{1-\mu\sum_i \alpha_i}} \kappa_k^{-1} L_k \tag{2.B13}$$

The eq. (2.B13) can be plugged into the regional aggregated good production[25]: $Y_c = A^Y_{a_c} \theta^{1-\sum_i \alpha_i} H_c^{1-\sum_i \alpha_i} \prod_i L_{c,i}^{\alpha_i}$ where $\theta = \frac{1-\sum_i \alpha_i}{1-\sum_i \alpha_i + \frac{1-\mu}{\mu}}$ is the share of regional land allocated to the production of the aggregated good. Doing so, one obtains the following regional real output expression:

$$Y_c = \theta^{1-\sum_i \alpha_i} (A^Y_c(a_c))^{\frac{1}{1-\mu\sum_i \alpha_i}} H_c \left(\prod_i A^{U_k}_c(a_c)^{\frac{\alpha_i}{1-\mu\sum_i \alpha_i}}\right) \prod_i \kappa_i^{\alpha_i} \tag{2.B14}$$

We now have an expression for both regional production (eq. (2.B14)) and regional labor demands (eq. (2.B13)) that can be plugged in eq. (2.B10). This characterizes the K wage

[24]Note that using the same assumptions, I can retrieve the extended wage and rent equations of Ottaviano and Peri (2006) as clearly defined in Bellini et al. (2013), p.14. The wage equation is similar to the one presented in this section, but it contains the indirect utility term, hence the choice of these further developments.

[25]This expression is easily obtained through the previous developments while summing the output of local homogeneous firms.

equilibrium equations in our model as well as the rent equation:

$$w_{c,k} = \theta^{1-\sum_i \alpha_i} \frac{\prod_i \kappa_i^{\alpha_i}}{\kappa_k} \frac{\alpha_k}{A_c^{U_k}(a_c)} A_c^Y(a_c)^{\frac{1-\mu}{1-\mu\sum_i \alpha_i}} \left(\prod_i A_c^{U_i}(a_c)^{\alpha_i} \right)^{\frac{1-\mu}{1-\mu\sum_i \alpha_i}} \tag{2.B15}$$

Several things emerge from eq. (2.B15); first, the presence of regional features for which a worker has a taste is negatively associated with wages, as workers are willing to accept a lower wage to take advantage of local amenities (and vice versa). Second, since these same characteristics may have a direct positive or negative effect on wages, one must consider which effect dominates the other. Finally, the novelty brought by the heterogeneity of workers with respect to its ancestor model comes from the last terms reflecting the preferences of other classes of workers: As these terms are derived from local labour demands (2.B11), it does mean that a region with characteristics that attract other workers will be associated to higher wages for a group of workers (and vice-versa). By further developments, one obtains the rent equation:

$$r_c = G((A_c^Y(a_c))^{\frac{1}{1-\mu\sum_i \alpha_i}}) \prod_i (A_c^{U_i}(a_c))^{\frac{\alpha_i}{1-\mu\sum_i \alpha_i}} \tag{2.B16}$$

Where $G = \Lambda^{\frac{1}{1-\mu\sum_i \alpha_i}} B^{\frac{\sum_i \alpha_i}{1-\mu\sum_i \alpha_i}} (\prod_i v_i^{\frac{\alpha_i}{1-\mu\sum_i \alpha_i}\alpha_i})$ is a constant. Together, the wage and rent equations capture the equilibrium relationship between local amenities and factor prices. In equilibrium, welfare is the same across locations, but regions that are more productive or have desirable local endowments have higher nominal wages, larger populations, and also higher housing costs.

2.B.6 GDP estimates appendix

Article	indep.	dep.	scale	level	start	end	OLS	IV
Docquier et al. (2020)	Migr. div. (bpl)	GDP pc (log)	USA	Commuting Zones	1960	2010	0.275***	1.063***
Bellini et al. (2013)	Migr. div. (citiz.)	GDP pc (log)	Europe	NUTS3 Reg.	1991	2001	0.950***	0.123
Ager and Brückner (2013)	Tot. div. (bpl)	GDP pc (log)	USA	Counties	1870	1920	-0.4**	-0.6**
Di Berardino et al. (2021)	Migr. div. (bpl)	GDP pc (growth)	Italy	Provinces	2003	2017	0.0738*	0.209**
Limonov and Nesena (2016)	Tot. div. (bpl)	GDP pc (avg. growth)	Russia	Reg. (Adm.)	2002	2011	0.089**	
Roupakias and Dimou (2020)	Migr. div. (bpl)	GDP per em- ployee (log)	Greece	NUTS2 Reg.	1991	2011	-0.36*	2.473
Minale et al. (2024)	Past Migr. div. (bpl)	GDP pc (log)	Brazil (SP)	Municipalities	1920	2000	0.285*	0.755
Alguacil and Alamá-Sabater (2021)	Migr. div. (en- tropy)	GDP pc (log)	Spain	NUTS3 Reg.	2002	2015	0.099	
Rodríguez-Pose and Von Berlepsch (2019)	Past Tot. div. (bpl)	GDP pc (log)	USA	Counties	1880	2010	0.144	0.371
Dohse and Gold (2014)	Tot. div. (citiz., fract.)	GDP pc	EU27	NUTS2 Reg.	2005	2010	0.031	
Dohse and Gold (2014)	Tot. div. (citiz., Theil)	GDP pc	EU27	NUTS2 Reg.	2005	2010	0.062	
Brunow and Brenzel (2011)	Migr. div. (bpl)	GDP pc (log)	Europe	NUTS2 Reg.	2004	2008	0.0004	0.004

Note: The table reports OLS and IV estimates from the economic literature studying the relationship between diversity and regional economic performance. Migr. div. refers to diversity among migrants, Tot. div. to diversity in the total population. By default, diversity is measured using a fractionalization index unless otherwise stated (entropy or Theil). GDP pc denotes GDP per capita.

Table 2.B13: Literature Estimates - Diversity and Gross Domestic Product

2.B.7 Instrumental strategy checks

This section tests the plausibility of the exogeneity assumption underlying the shift-share instrument employed in table 2.2, following the methodology advanced by Borusyak et al. (2022). The instrumental strategy relies on the exogeneity of the national shocks to metropolitan wages. This assumption is more plausible when shocks are numerous and not concentrated in a small number of birthplaces. To test these assumptions, the table 2.B14 presents descriptive statistics of the national migrant birthplace population growth rates (from 2000) and the inverse Herfindahl Index of the migrant birthplace shares. This last measure is used to evaluate whether the distribution of migrant birthplaces in the U.S. is sufficiently diffuse and whether individual birthplaces contribute relatively small shares to the national total. The descriptive statistics are weighted according to the lagged national birthplace shares to retrieve US representative estimates. As the section 2.B.1 fig. 2.B7 shows Mexican migrants represent about a quarter of the total US migration flows on average, I also compute the statistics while excluding Mexican origin in columns (2) and (5). The columns (3) and (6) present results using the 2015 and 2019 samples only. The inverse Herfindahl indices are lower when specific birthplaces have a strong weight among the workers' groups. They are high for the Industry-occupation groups level as expected from strongly disaggregated groups, and the results suggest the Mexican birthplace indeed has a relatively strong weight at the wider Experience-Education cell; the inverse Herfindahl index remains however relatively high with a value close to Borusyak et al. (2022) when they compute it with Autor et al. (2013)'s data.

An additional condition is that the shocks are sufficiently mutually uncorrelated. The intra-class correlation coefficients of shocks in the different origins and groups are computed from a random effects model considering aggregated sector or aggregated occupational groups and sectors (columns (1) to (3)) and diploma group or 5-years experience bins (columns (4) to (6)). The correlation estimates are all fairly low. I replicate the main regressions while

	Industry-occupations			Experience-education cells		
	(1)	(2)	(3)	(4)	(5)	(6)
Descriptive Statistics						
	All	No Mexican	2015-2019	All	No Mexican	2015-2019
Mean	0.548	0.665	0.633	0.322	0.348	0.404
Standard Deviation	1.667	1.949	1.877	0.954	1.008	1.205
Interquartile Range$_{75\%-25\%}$	0.916	1.150	1.035	0.778	0.811	1.148
$1/HHI$	1,241.77	12,804.09	1,241.77	192.95	2,356.19	192.95
Intra-Cluster Correlations						
Aggr. Industry/Diploma	0.0009	0.0002	0.0019	0.0314	0.0100	0.0333
Aggr. IndustryxOccupation/experience	0.009	0.0161	0.0090	0.0310	0.0515	0.0315

Note: This table summarises the distribution of the immigrant industry-occupational and experience-education cell birthplace migration growth rates since 2000 at the national level (metropolitan migrations only). 1/HHI is the inverse Herfindahl index of the initial birthplace shares. Workers groups are either defined as industry-occupational groups or as Ottaviano and Peri (2012) skill cells. The inter-cluster correlations are computed using the estimates of a random effect model following the Borusyak et al. (2022) methodology. The data are from the US IPUMS 2000-2019 databases. All descriptive statistics are weighted by the initial birthplace shares.

Table 2.B14: IV Shocks Summary Statistics and Inter-Cluster Correlations

excluding Mexican workers from migrant shares and migrant diversity indexes in table 2.B15. The results remain similar in size but the IV estimates are higher in absolute value and are not significant at the 10% threshold.

	Industry-Occupation			Education-Experience Skill cells		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Low-skilled	High-skilled	All	Low-skilled	High-skilled
Panel A: OLS estimates						
Migrant diversity$_j$	0.00923***	0.00808*	0.00469	-0.0121**	0.00290	-0.0179**
	(2.65)	(1.77)	(1.20)	(-2.09)	(0.44)	(-2.43)
Migrant share$_j$	-0.0215	-0.0202*	-0.00923	0.156***	0.111***	0.246***
	(-1.13)	(-1.82)	(-0.50)	(5.08)	(4.18)	(5.35)
$ln(Population)$	0.0565*	0.0395	0.0751**	0.0539	0.0555	0.0750*
	(1.74)	(1.11)	(2.15)	(1.48)	(1.42)	(1.86)
$ln(Population_j)$	0.00892***	0.00460***	0.0119***	0.0137**	-0.0148***	0.0103*
	(3.52)	(2.90)	(3.92)	(2.41)	(-3.14)	(1.81)
Panel B: IV estimates						
Migrant diversity$_j$	0.01725	0.01459	0.00927	-0.00305	0.00339	-0.0202
	(1.53)	(1.50)	(0.66)	(-0.22)	(0.20)	(-1.16)
Migrant share$_j$	-0.119***	-0.0916***	-0.0810	0.172***	0.0822***	0.411***
	(-2.84)	(-3.84)	(-1.59)	(5.28)	(2.82)	(7.37)
$ln(Population)$	0.0612*	0.0389	0.0812**	0.0558	0.0580	0.0682
	(1.74)	(1.02)	(2.13)	(1.50)	(1.44)	(1.64)
$ln(Population_j)$	0.0140***	0.00698***	0.0179***	0.0135**	-0.0158***	0.0111
	(4.65)	(3.63)	(5.04)	(2.15)	(-2.99)	(1.63)
Time FE γ_t	Yes	Yes	Yes	Yes	Yes	Yes
Skill Cell FE γ_j	Yes	Yes	Yes	Yes	Yes	Yes
MSA FE γ_c	Yes	Yes	Yes	Yes	Yes	Yes
Individual char. $X_{i,t}$	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,048,925	1,137,135	2,155,038	3,477,385	1,315,610	2,161,775
R-squared	0.409	0.307	0.402	0.317	0.210	0.283
K-P F-Stat	472.5	332.3	550.2	363.2	315.2	291.6

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, t statistics in parentheses. Robust se clustered at MSA-level. OLS Panel is a reproduction of table 2.B3 regressions by skill groups. Migrant diversity indexes and shares are computed at the metropolitan worker groups j level. Workers groups are either defined as industry-occupational groups or as Ottaviano and Peri (2012) skill cells. IV panel contains estimates using 2SLS estimates (Migrant share and diversity are instrumented using a shift-share strategy). Dependent variable is log of native $1999 hourly wage. X_{it} is a vector of individual characteristics containing age, age squared, racial and gender dummies. Source: Author's elaboration on IPUMS 2000-2019 ACS data.

Table 2.B15: OLS and 2SLS estimates by skill groups - Diversity indices and migrant shares computed excluding Mexican birthplace

Chapter 3

Migration, labor markets, and wage adjustment mechanisms: A shift-share approach

While an academic consensus states that average native wage and employment do not respond to migration shocks in the long run, as capital adjusts to changes in the labor force, empirical estimates of the short-run responses following migration episodes appear to be heterogeneous, particularly between experimental and quasi-experimental approaches. This paper argues that much of this divergence stems from methodological choices and investigates, in particular, the role of the shift-share instrumental strategy. Using a harmonized U.S. dataset, it shows that estimated migration effects on native wages are sensitive to the level of analysis and to how the instrument is implemented. By systematically varying these features, the paper replicates the heterogeneity of results documented in the literature within a single empirical framework, and reconciles spatial correlation methods with experimental findings: both coexist once the cumulative nature of migration shocks is accounted for. The analysis further evaluates the performance of shift-share instruments across skill and geographic levels of aggregation, revealing a fundamental trade-off between the ability to capture native spatial and occupational sorting and the exogeneity of the instrument.

3.1 Introduction

Immigration to high-income countries increased greatly in the last few decades, polarizing the public debate, especially regarding its impact on the labor market of the host states. An abundant literature studies the impacts of migrations on economic outcomes in destination countries (Stark and Bloom, 1985; Polachek et al., 2006; Borjas, 2018) with various strategies. An academic consensus states that the average native wage does not respond or responds only weakly to migration shocks in the long run as capital adjusts to changes in the labor force (Nedoncelle et al., 2025; Edo, 2019). However, the short-run response of wages following migration episodes appears to be highly heterogeneous (Dustmann et al., 2016).

Part of this heterogeneity likely reflects differences in methodology. The most common reduced-form approaches include *spatial correlation studies* (Card, 2001; Zorlu and Hartog, 2005; Jaeger et al., 2018), which compare geographic units with different levels of migration exposure; *skill-cell studies* (Borjas, 2003; Ottaviano and Peri, 2012; Caiumi and Peri, 2024), which exploit variation across national skill groups; *mixed approaches* combining both dimensions; and *experimental methods* (Card, 1990; Cohen-Goldner and Paserman, 2011; Glitz, 2012), which exploit sudden and large migration episodes. Despite broadly similar contexts, these approaches tend to produce divergent results. Estimated effects vary across levels of geographic and occupational aggregation, which may reflect genuinely different adjustment processes, but also endogeneity concerns (Jaeger et al., 2018). While experimental studies are less exposed to income pull-factor bias, their external validity may be limited, as they capture sudden and massive shocks that are not necessarily representative of regular migration flows. As the influence of standard migration patterns on economic outcomes cannot be identified through experimental methods, quasi-experimental settings are required. In particular, an increasing share of contributions across all these families relies on the shift-share instrumental strategy (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022), which exploits pre-existing diasporas' settlement patterns to construct plausibly exogenous variation in immigration exposure.

This chapter contributes to the literature by providing a systematic, harmonized reassessment of the labor market effects of immigration using a common US database covering a timeline from 1980 to 2019, and a consistent empirical framework across most of the levels of analysis traditionally considered, covering both *skill cell* and *spatial correlation* strands as well as their intersections. The existing literature reports a wide range of estimates based on different datasets, geographic units, sample definitions, and identification strategies, making direct comparisons hazardous. By replicating the main approaches from the literature with a unified dataset, this research isolates the extent to which divergent findings reflect methodological choices rather than underlying economic contexts. It also evaluates the performance of shift-share instruments in migration economics and tests recent criticisms related to the dynamic nature of migration effects (Jaeger et al., 2018; Monras, 2020).

The results suggest a weak performance of the shift-share strategy for skill-cell studies. Additionally, it pictures how the shift-share instrumental strategy, as it is commonly used in the migration literature, is responsible for an omitted variable bias due to serial correlation (Jaeger et al., 2018), which makes the numerous contributions using this identification tool potentially misguided. Using an older version of the shift-share instrument that captures the cumulated long-run effect of all migration shocks supposed, it provides evidence that a single migration shock with migrants representative of regular migration flows had an overall negative impact on average. This result reconciles spatial correlation methods with findings exploiting sudden migration shocks that do not suffer from serial correlations. Finally, if the estimated bias sign was common to all levels where the instrument had a decent strength, the results still show potential positive short-run effects at the industrial level that would require further research.

The chapter is structured as follows: The second section provides a summary of the state of our knowledge on the impact of migration on host countries' labor market, as well as general considerations. The third section presents the data, the general theoretical considerations behind the shift-share empirical strategy used for this research, and the general structure guiding the empirical analysis. The fourth section introduces the shift-share instrumental approach as well as its main form in migration studies, while performing different checks on the main assumptions the strategy builds on. A fifth section presents the main results, followed by a discussion on the general findings and comments on further possible research developments.

3.2 Literature review

3.2.1 The short and long-run dynamics of the labor market in response to migration shocks: A general overview

The literature studying the impact of migration on wages and employment emphasizes both short-run and long-run effects. Firstly, migration shocks are associated with a direct increase in the labor supply, resulting in lower mean native wages if the degree of substitution between natives and immigrant workers is high. Their relative skill composition is key to this parameter: If they are similar, the average wage will decrease in the short run following a migration shock, while sizable differences in skills will result in a native wage increase because of their complementarities (Dustmann et al., 2016, pp.40-41). Behind the general effect, various findings suggest the coexistence of negative and positive influences of migration on wages for different groups of workers (Manacorda et al., 2012; Ottaviano and Peri, 2012; Dustmann and Preston, 2012; Edo and Toubal, 2015). As patterns of migrations are relatively stable across time, migrations typically pressure previous migrants' wages as they tend to share the same skills (Ottaviano and Peri, 2012; d'Amuri et al., 2010). Similarly, if migrants are more educated on average than the native population, native educated workers will face a decrease in their income, while other workers will benefit from the changes in the skill composition of the country. It then decreases the inequalities between these two segments of the population and has a positive short-run impact on natives' wages (Edo and Toubal, 2015).

In the longer run, the mobility of capital and workers adjusts to the new labor force and composition, and the average marginal productivity of aggregate labor recovers its past level (Lewis and Peri, 2015; Amior and Manning, 2018; Edo and Toubal, 2015; d'Amuri et al., 2010). This null influence is remarkably stable across empirical evidence (Edo, 2019), with the exception of some works considering a very long time horizon, and finding that strong migration flows contributed to long-run growth through innovation or early agglomeration economies (Sequeira et al., 2020; Peters, 2022). Similar to a capital adjustment mechanism, it could reflect long-run dynamics related to a boost in the demand for houses (Howard, 2020), resulting in new construction activities. Other research also demonstrates a positive influence of high-skill migrant presence on trade (Aleksynska and Peri, 2014), potentially increasing production in exporting sectors, and on innovations (Bosetti et al., 2015; Hunt and Gauthier-Loiselle, 2010) that may increase workers' productivity in the long run. In the presence of internal labor mobility rigidities, agglomeration economies may also result in an increase in productivity in the long run, supported by capital flows and cross-fertilization of ideas. Focusing on the cycle of capital adjustment in wage recovery, as structural and shock-exploiting studies typically do, may exclude these dynamics from overall measured effects: Exploiting an anterior massive flow of ethnic Germans expelled from their homes in Eastern Europe after the Second World War, Braun and Weber (2021) as well as Braun and Mahmoud (2014) find a strong adverse negative effect on labor market outcomes. Exploiting exactly the same shock but looking at its influence decades after, Peters (2022) finds that the subsequent refugees' settlements resulted in an industrialization of exposed rural areas, boosting the local productivity in the long run. Building a counterfactual from structural estimates, a key element driving his results is the presence of spatial frictions limiting labor mobility with respect to capital. While the results of structural studies may also reflect modeling assumptions, they show that migration tends to have dynamic effects on economic outcomes.

The duration of the recovery from an immigration shock across skill groups theoretically depends on differences in the shock exposure, but it may also be rooted in the differences of capital complementarities of occupations under the argument of Lewis and Peri (2015): If capital adjusts for supply shocks in specific jobs, it will be faster if non-exposed jobs can easily be substituted by machines to absorb new workers in exposed jobs. The skill composition should therefore also influence the persistence of the effect of immigration on labor market.

The reduced-form empirical studies of the short-run responses of wages to migration are, however, more divided both in terms of results and methods. The next sections briefly introduce these approaches and present a large subsample of their estimates, supposed to be representative, to retrieve a general overview of the findings categorized by their methodology. It also exposes potential limitations behind these different patterns.

3.2.2 Spatial correlation studies

A common approach to assess the influence of migration exposure on regional economic outcomes is to compare regions more exposed to migration with less exposed regions. This approach is particularly convenient for the use of an instrumental strategy based on network spatial distribution, such as the Bartik/shift-share instrument, hence its popularity in the subsample of spatial correlation studies depicted in table 3.1. With a bit of variation, this part of the literature emphasizes null or slightly positive effect of migrations in the sort-run.

A first limitation of this approach is its relative dependence on the shift-share instrumental strategy that may cause this apparent homogeneity due to a bias toward 0 related to the serial correlation of migration flows, as discussed by Jaeger et al. (2018) and Monras (2020). This issue will be discussed in section 3.4.

A second limitation of this approach lies in the fact that it fails to take into account spatial redistribution of workers in response to migration shocks. About half of interstate migrations are indeed driven by work opportunities (Shambaugh et al., 2017, p.7). Natives may respond to spatial wages differences (Borjas and Edo, 2021; Amior and Manning, 2018) and higher housing costs (Howard, 2020) by moving to markets that are either targeted by migration shocks or the contrary. Colas (2019) estimates that after ten years, half of the US local wage impact from a migration shock dissipated through this channel. Estimating the impact of the influx of ethnic Germans in specific Western German regions following the fall of the *Berlin* wall, Glitz (2012) documents that this influx resulted in internal movements of native workers: For ten migrants arriving in a receiving region, about 3 local workers left. This mobility (as well as German institutional rigidities) may explain the absence of impact found in this study. Aggregate spatial levels helps to mitigate such concerns.

3.2.3 skill-cell studies

The term *skill-cell studies* usually refers to research evaluating the effects of migration at the level of specific national worker groups. By centering their analysis at the most aggregate geographic level, they mechanically retrieve national estimates capturing geographic sorting.

This approach tends to give more weight to competition as it assesses the influence of migration shocks on the wages of native workers sharing their skill group. For instance, if the level is occupation, it does not directly take into account that local workers tend to change jobs in response to migration shocks. For instance, Foged and Peri (2016) exploit a national policy of refugees' allocation in specific regions to assess how this positive shock in the labour supply actually increased the wages and employment of skill-similar natives, showing that unskilled workers self-select into less manual-intensive occupations in response. These results are consistent with the existence of the short-run occupational arduousness gap between natives and immigrants in Europe described by Vandenberghe (2025), indicating that the skill cell approach may lack external validity. However, one can play on the definition of skill cell to modify the ability to capture such effects. To retrieve a more stable definition of a skill cell that would capture occupational sorting, they are often defined as education-experience bins based on the level of diploma and intervals of years of experience, following the seminal work of Borjas (2003). The main idea is to define groups of workers such that a single cell's members are relatively similar in skill but are unable to sort into another cell. If workers' complementarities pass through occupational adjustments, this approach allows for capturing them. However, in the presence of *downgrading*, the tendency of migrants to sort into jobs in lower positions on the labor market than natives with similar qualifications (Zorlu, 2016), this approach may miss important elements. Notably documented by Dustmann et al. (2012) for the UK and Dustmann and Preston (2012) for the US, this feature of new migrants tends to dissipate with years, as they adapt their linguistic or other skills relevant to their new environment. In Europe, Vandenberghe (2025) typically shows that first-generation migrants self-select in jobs with more arduous tasks than natives, but that this tendency fades progressively years after arrival. In the presence of serial correlation, it could result in an ambiguous bias summing these changes.

With a stable skill cell definition and such consideration, this literature emphasize mostly negative and null results in table 3.2, but these estimates may be affected by the serial correlation mentioned earlier in a different way than spatial correlation studies.

3.2.4 Experimental studies

A strand of the literature exploits specific unexpected migration episodes as natural experiences. These unexpected, often massive flows are likely to be exogenous to local wealth as they are often made feasible by a supposedly unrelated event: The fall of the *Berlin* wall (Glitz, 2012) or the repatriation following the independence of a colony (Edo, 2020), for example. If the internal validity of these works is often difficult to question, they tend to represent a rather local effect, as these shocks are not likely to reflect traditional economic migrations. A subsample of the most influential of them is available in table 3.3. Potentially because of the importance of such scarce shocks, their impact on native wages tends to be negative and relatively important. An interesting feature of these shocks is their absence of serial correlations, allowing to directly assess the dynamic response of economic variables: Studying the impact of the sudden massive Jewish migration shock from the USSR to Israel in 1990, Cohen-Goldner and Paserman (2011) find a negative impact on the average occupational wage that took 5 to 7 years to dissipate, according to the skill. Exploiting French repatriates following the independence of Algeria in 1962, Edo (2020) shows a negative influence on both wage and employment on regional native wages, that lasted for more than 10 years.

Country	Period	Study	Group of natives	Level	Id. strategy	Sign	Elast.	Emp.
Australia	2011	Breunig et al. (2017)	Men and women	Region		0		0
Austria	1981-1991	Winter-Ebmer and Zweimüller (1996)	Young blue-collar workers	Region	IV: Lagged migrations	0		
	1988-1991	Winter-Ebmer and Zweimüller (1999)	Young men	Region	IV: structure of employment & lagged migr.			$-$
France	1996-2005	Mitaritonna et al. (2017)	Men and women	Department	IV: Shift-share	$+$	0.50	
	1976-2007	Ortega and Verdugo (2016)	Male blue-collar workers (2555)	Emp. zone	IV: Shift-share	$-$	-0.10	
	1976-2007	Ortega and Verdugo (2022)	Male blue-collar workers (2555)	Commuting zone	IV: Shift-share	$-$	-0.33	
Germany	1985-1989	Pischke and Velling (1997)	Men and women	Region	Conditioning: Past labor market out.			0
Great Britain	1983-2000	Dustmann and Fabbri (2005)	Men and women	Region	IV: Shift-share	0		0
The Netherlands	1997	Zorlu and Hartog (2005)	Men and women	Region	IV: structure of employment & lagged migr.	≈ 0		
Norway	1996	Zorlu and Hartog (2005)	Men and women	Region	IV: structure of employment & lagged migr.	≈ 0		
United Kingdom	1997-2005	Dustmann et al. (2012)	Men and women	Region	IV: Shift-share	$+$	0.26	
	1997	Zorlu and Hartog (2005)	White men and women	Region	IV: structure of employment & lagged migr.	≈ 0		
United States	1970-1980	Altonji and Card (1991)	Low-skilled	Metro area	IV: Shift-share	$-$	-1.20	0
	1980-2010	Basso and Peri (2015)	Men and women	State	IV: Shift-share	0		
	1980-2000	Card (2007)	Men and women	City	IV: Lagged migrations	$+$		
	1970-1980	Jaeger et al. (2018)	Men and women	Metro area	IV: Shift-share	$-$	-0.43	

Notes: $+$ denotes a positive and statistically significant wage effect, $-$ a negative and statistically significant effect, 0 a statistically insignificant effect, and ≈ 0 a negligible effect. "Est.." reports the wage elasticity with respect to immigration when available. Autor's extension of the original Table 3 of Edo (2019).

Table 3.1: Selected Studies from Spatial Estimations

Country	Period	Study	Group of natives	Unit of analysis	Sign	Est.	Emp.
Australia	2003-2012	Breunig et al. (2017)	Men and women	Education-experience	0		0
Canada	1980-2000	Aydemir and Borjas (2007)	Men	Education-experience	–	−0.32	
France	1990-2002	Edo and Toubal (2015)	Men	Education-experience	0		–
	1990-2002	Edo (2016)	Men on fixed-term contracts	Education-experience	–	−0.79	0
	1968-1999	Ortega and Verdugo (2014)	Men	Education-experience	+	0.33	+
Germany	1975-1997	Bonin (2005)	Men	Education-experience	–	−0.10	
	1992-2001	D'Amuri and Peri (2014)	Men and women	Education-experience			0
	1975-2001	Steinhardt (2009)	Men	Occupation	–	−0.16	
Norway	1998-2005	Bratsberg and Raaum (2012)	Men and women in construction sector	Task	–	−0.44	
Spain	2002	Carrasco et al. (2008)	Men and women	Education-experience	0		
	1991-2001	Carrasco et al. (2008)	Men and women	Education-experience			0
United States	1960-2000	Borjas (2003)	Men	Education-experience	–	−0.40	
	1960-2000	Borjas et al. (2010)	White men	Education-experience	–	−0.39	–
	1960-2010	Borjas (2014)	Men and women	Education-experience	–	−0.25	
	1970-2000	Llull (2018)	Men	Education-experience	–	−1.20	

Notes: + denotes a positive and statistically significant wage effect, − a negative and statistically significant effect, 0 a statistically insignificant effect. "Elast." reports the wage elasticity with respect to immigration when available. "Emp." reports the employment outcomes when available.

Table 3.2: Selected skill-cell studies

Country	Episode	Study	Sample	Level	Wage	Est.	Emp.
European countries	Refugee flows from Yugoslavia (19911992)	Angrist and Kugler (2003)	Men and women	Country	–		
France	Algerian Rapatriates	Hunt (1992)	Men and women (1962-1968)	Department	–	-0.80	–
		Edo (2020)	Men (1962-1969)	Region	–	[-1.0; -2.0]	–
		Edo (2020)	Men (1962-1976)	Region	0		
Germany	Fall of Berlin Wall inflows	Glitz (2012)	Men and women	Region-occupation	0		–
	Czech commuting policy (1991)	Dustmann et al. (2017)	Men and women	Municipality	–	-0.13	–
Portugal	Repatriation from colonies (1970s)	Carrington and De Lima (1996)	Portuguese	Country	0		–
		Mäkelä (2017)	Portuguese	Country	–		–
Israel	USSR immigration wave (1990)	Friedberg (2001)	Men and women	Occupation	0		0
		Cohen-Goldner and Paserman (2011)	Men and women	Occupation	–	[-0.1; -0.3]	0
	47 years after	Cohen-Goldner and Paserman (2011)	Men and women	Occupation	0		0
Turkey	Syrian refugees (2012)	Tumen (2016)	Men and women	Region	0		–
United States	Mariel Boatlift (1980)	Card (1990)	Men and women	City	0		0
		Borjas (2017)	Low-skilled non-Hispanic men	City	–	[-0.5; -1.5]	
		Peri and Yasenov (2019)	Men and women	City	0		0
		Borjas and Monras (2017)	Non-Hispanic men	Metro area (education)	–	-1.3	0
	Mexican peso crisis	Monras (2020)	Men and women	Metro area (education)	–	[-0.7; -1.4]	–

Notes: + positive effect; − negative effect; 0 no statistically significant effect. “Est.” reports estimated wage effects range when available. Author’s extension of Table 5 of Edo (2019).

Table 3.3: Selected natural experiment studies

3.3 Data and framework

3.3.1 Data

To investigate the relationship between wages or employment share and migrations, I exploit the Integrated Public Use Microdata Series (IPUMS) USA datasets (Ruggles et al., 2025) listed in section 3.B.1. More precisely, the core of the analysis focuses on various regional and skill cell levels across four periods: 1990, 2000, 2007 and 2019. The constraints behind this specific choice and the (slightly) unbalanced periods are variable availability and the will to limit contamination from the COVID-19 pandemic in the estimates. For robustness checks, more and less temporally distant sample periods (2005, 2010, 2015, and 2019 samples or 1990, 2006, and 2023 samples) are also used. IPUMS sample weights are applied to aggregate individual information at geographic and skill cell levels, in order to retrieve representative values. Variables expressed in monetary units are adjusted for inflation using the Consumer Price Index (ref. \$ 1999) of the Bureau of Labor Statistics, directly available in the datasets. IPUMS Censuses report individual yearly wage income, the number of weeks worked, and hours worked by week. These variables allow for computing average and median hourly wages for each region and workers' category. The samples also contain information on the birthplace and other demographic characteristics necessary to compute migration exposure at the desired levels of analysis. Employment share is computed as the share of native in the labor force among the total native age-working population. This measure can be obtained for occupational and industrial groups as IPUMS data contains the previous year's occupational information on individuals; this measure then reflects the number of individuals staying in the occupation or the industry, divided by a specific age-working population, including them and the workers who lost their job/resigned. It is therefore different from the employment rates measured at other levels due to its short-run definition. Descriptive statistics of these variables are available for different levels of analysis and different samples in section 3.B.2 table 3.B1 to table 3.B6.

3.3.2 Methodology

Based on similar theoretical considerations to the ones covered above, different approaches exist to quantify the impact of immigration on labor market outcomes (Dustmann et al., 2016; Edo, 2019). Some exploit the difference in migration exposures between geographic areas, others between skill groups, or both. The following baseline summarizes the joint approach applied in this research:

$$ln(w_{cjt}) = \delta_0 + \delta_1 Migrants\ share_{cjt} + Z'_{cjt}\Gamma + \gamma_j + \gamma_c + \gamma_t + \epsilon_{cjt} \tag{3.1}$$

Where w_{cjt} is the average hourly native wages of workers in skill group j in area c at period t. $Migrants\ share_{cjt}$ is the ratio of the migrants and the total workers' population. If the native population is fixed, its increase represents a migration-induced labor supply positive shift. Z'_{cjt} is a vector of observable adjustment factors. In most specifications, it only includes an adjustment term for the incomplete weight of the instrumental strategy described in the following section, but additional checks add a lagged migration exposure to account for the past adjustment process. The terms γ_j, γ_c, and γ_t represent, respectively, skill group, area, and period fixed effects, while ϵ_{cjt} is an error term. The standard errors are clustered at the interacted skill group - area level. This equation is estimated for different groups j at different levels of geographic aggregation c. This heterogeneity in levels of analysis has three main advantages: First, it allows for comparing the results directly to most of the related literature. Doing so, it provides evidence that the same opposite general conclusions can be drawn from the same dataset, highlighting the importance of the level of analysis in migration studies' outputs. Second, it allows for exploring the influence of native mobility in the overall effect of migrations by comparing the results from aggregated levels to local ones: Broad geographic area average wages take into account geographic sorting of workers in response to migration shocks, while small areas' native wages reflect a mixed influence on workers with less mobility and internal migrants complementary to external migrants. Finally, it allows for investigating the properties of the estimates at different levels of aggregation. Note that a particular feature of some skill-cell studies is to include an interacted skill-time fixed effect to absorb differences in time trend between skill groups. As it makes the source of variations behind the estimates difficult to interpret (Dustmann et al., 2016), originating from a triple difference, which would make comparisons more perilous, I stuck to the eq. (3.12) nevertheless. I define the variables at four main geographic levels: National, Region, State, and county. Under the same reasoning, the definition of the workers' groups matters to capture the switch of jobs or industries. I consider five levels: the broadest are the global level with all workers and the skill cell following the approach of Ottaviano and Peri (2012). It consists of dividing workers according to their years of education and experience. As these authors, I use six educational groups of workers defined by attainment: High school dropouts, High school graduates, Some college education, Bachelor's, Master's, and doctoral degree. I use the age of workers and their theoretical years of studies to retrieve a measure of potential professional experiences, and create bins of 5-year intervals. The interaction of educational and experience bins gives the skill cells. For example, a cell is the workers with a Master's education and 25 to 29 years of potential experience. I also define skill as the industry of the occupation, the occupation itself, and their interaction (occupation in an industry). These last levels gather workers with highly similar skills and tasks, while occupational mobility can occur at the industrial level. Industrial mobility could occur at the occupational level. Their intersection allows for considering workers who did not move to another occupation or industry.

3.4 Identification and shift-share instrumental approach

3.4.1 Main idea

The main threat for causal identification arises from the fact that migrations are not random and are typically driven towards regions with higher wages (Peri, 2016). To avoid capturing the induced circular relationship as well as potential confounding factors, an entire strand of the literature uses a *shift-share* instrumental strategy based on section 3.4.2 theoretical consideration. This instrument is an interaction of national inflows by country of origin with the spatial and/or skill historical distribution of the diaspora of the same origins (Card, 2001, 2009). It relies on the idea that migrants are attracted by the presence of persons of the same origins (McKenzie and Rapoport, 2010; Beine et al., 2011). Its original form is:

$$z_{ct} = \sum_{o=1}^{O} \underbrace{\frac{M_{co,t_0}}{L_{c,t_0}}}_{s_{co}} \underbrace{(\frac{M_{ot} - M_{o,t_0}}{M_{o,t_0}})}_{g_{ot}} \tag{3.2}$$

where g_{ot} represents the migration growh rate of origin o between the reference period t_0 and the period t. The term s_{co} is the share of the origin o immigrant population in region c at the reference period. Their products represent naive predictions of the local labor increase of workers born in each origin o, assuming the spatial distribution of these new migrants is fully explained by the historical diaspora. Their sum across origins, z_{ct}, is therefore the total predicted immigration-induced regional labor growth rate. It is supposed to capture variations rooted in diaspora *pull-factors* that would be exogenous to local wages. This independence notably implies that the presence of migrant communities during the reference period does not affect future economic outcomes. The further in the past the concerned diasporas, the more credible this assertion is, but the weaker the instrument. However, it also builds on the assumption of an absence of long-run migration influence, notably from bigger migrant communities. If one wants to remain agnostic with respect to this idea, the exogeneity remains valid only under the condition of the exogeneity of the national inflows to the wages of the workers' groups (Borusyak et al., 2022).

This strategy has therefore two costs: While often used in the literature at highly aggregated (often national skill cell) levels, this second condition implies that the wages at the level of the study are exogenous to the national migration shocks. This condition is likely to hold at the disaggregated levels, but is more difficult to defend for broader groups. For instance, if specific groups of migrants have a strong weight in the instrument, one needs to be aware of their specific characteristics that may influence wages in the long run. This research provides the opportunity to check if the conditions hold at any level traditionally considered, using the tests proposed by Borusyak et al. (2022). The next section shows a theoretical consideration behind this instrumental approach.

Moreover, shift-share predicted migrant populations change may be correlated with past shock of migration still influencing wage contemporary. In the absence of unexpected shocks, observed migration flows to a country tend indeed to be serially correlated. As the past diasporas' presence also attracted migrants at the previous period and if national shifts are serially correlated, it would induced a bias in the estimated short-run effect on wage as the instrument would include a weighted effect of all migratory shocks serially correlated with the national shift on contemporary wages (Jaeger et al., 2018). The following sections will discuss the theoretical foundation of the shit-share instrument as well as the consequences of such serial correlation for the interpretation according to two different shifts that can will be used.

3.4.2 A theoretical foundation

External migrations

Jaeger et al. (2018) propose the following theoretical foundation for the shift-share instrumental strategy widely used in migration studies: Migrants choose their regional location in the host country based on potential income and their diaspora presence. They maximize the following utility function:

$$U_{cjot} = U(\frac{M_{cjo,t_0}}{M_{o,t_0}} ; \frac{w_{cjt}}{\bar{w_{jt}}}) \tag{3.3}$$

With M_{cjo,t_0}, the number of migrants working in region c at the reference period t_0 sharing the origin o specialized in tasks j. The term M_{o,t_0} measures the same diaspora presence at the national level. Hence, their ratio represents the relative presence of origin o in the associated region at time t_0. w_{jt} is the real wage of a worker j in region c at time t, and $\bar{w_{jt}}$ is its national average. The marginal utility from both components is positive: relative income and the diaspora o are migration *pull-factors.* Considering an exogenous arrival of migrants from various origins in the destination country, its allocation across regions reads:

$$m_{cjt} \simeq (1-\lambda) \sum_{o} \frac{M_{ocj,t_0}}{M_{o,t_0}} \frac{\Delta M_{ot}}{L_{j,t_0}} + \lambda(\frac{w_{cjt}}{\bar{w_{jt}}} \frac{\Delta M_{ot}}{L_{j,t_0}}) \tag{3.4}$$

This equation postulates that the spatial distribution of migrants depends on the diaspora presence through the first term and on regional wealth in the second. The first term is the sum of the relative presence (share) of an origin in a region multiplied by the migration growth (shift) of the same origin to the host country. It captures the spatial distribution of migrations related to cultural pull factors. The parameter $1 - \lambda$ measures its relative importance with respect to income pull factors.

Production setting

Borrowing elements from Card (2001), each region of the host country produces a single output good at each period in quantity Y_{ct}, using a CRS Cobb-Douglas technology:

$$Y_{ct} = A_{ct} K_{ct}^{\alpha} L_{ct}^{1-\alpha} \tag{3.5}$$

Each worker supplies a fixed amount of work, normalized to 1, in her current occupation j. L_{ct} is a CES aggregator of the total labor supply in each occupation L_{cjt}:

$$L_{ct} = (\sum_j a_{cjt}(M_{cjt}) L_{cjt}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} \tag{3.6}$$

The term $a_{cj}(M_{cj})$ represents regional and occupation-specific productivity that can be affected by the migrant presence M_{cjt}. The term σ is the elasticity of substitution between labor factors. From these elements and assuming perfect competition, the (log) wage equation of a factor j reads:

$$ln(w_{cjt}) = \theta_{ct} + ln(a_{cjt}(M_{cjt})) - \frac{1}{\sigma} ln(L_{cjt}) \tag{3.7}$$

Where $\theta_{ct} = ln(A_{ct} K_{ct}^{\alpha} L_{ct}^{\frac{1-\alpha\sigma}{\sigma}})$ captures notably the influence of capital flows and the size of the local economy. When differentiating this equation and assuming native labor immobility in response to a migrant shock m_{cjt}, one obtains:

$$\Delta ln(w_{cjt}) \simeq \Delta\theta_{ct} + \Delta ln(a_{cjt}(m_{cjt})) - \frac{1}{\sigma} m_{cjt} \tag{3.8}$$

If workers and capital are perfectly immobile, an immigrant-induced labor supply shock has a net impact on local wages equal to $a'_{cjt} - \frac{1}{\sigma}$. If there is no investment friction, capital flows immediately compensate it through the term $\Delta\theta_{ct}$ due to the constant returns to scale. Indeed, if we define $k_{ct} = \frac{K_{ct}}{L_{ct}}$, the capital per unit of aggregated labor, we have:

$$ln(k_{cjt}) = \frac{1}{1-\alpha} ln(\frac{\alpha}{r}) + \frac{1}{1-\alpha} ln(A_{ct}) \tag{3.9}$$

Capital per aggregated labor unit depends on local productivity and the return to capital r, not on the level of aggregated labor. It therefore adapts to a shock by reaching the same level as previously. However, both perfect and impossible mobility of factors are not plausible hypotheses in the short run; relaxing them implies dynamic progressive adjustment of capital and labor.

3.4.3 Varieties of shift-share instrumental strategy: overall, short-run, and adjustment detections

The *original* version of the shift-share instrument presented in section section 3.4, that I will call the *short-run instrument* for reasons that will be discussed, can be written as in eq. (3.11). It has many variants[1]. While it imposes a reference period t_0 both for the shares and the shifts components of the instrument, it is common practice to slightly modify its formula by focusing on changes in predicted immigrant presence, using separate reference periods for the shares and the shifts. For instance, if one, following Card (2009) or Jaeger et al. (2018), considers the shift component as a specific increase in the number of migrants rather than a growth rate from the same reference as the shares, then one can write the predicted migrant share as the ratio between predicted migrant population change and the population at a previous period $t-1$ as in eq. (3.10):

$$\text{Short-run migr. instrument} \qquad z'_{ct} = \sum_{o=1}^{O} \underbrace{\frac{M_{co,t_0}}{M_{o,t_0}}}_{p_{co}} \underbrace{\left(\frac{M_{ot} - M_{o,t-1}}{L_{c,t-1}}\right)}_{\gamma_{ot}} \tag{3.10}$$

$$\text{Cumulated migr. instrument} \qquad z_{ct} = \sum_{o=1}^{O} \underbrace{\frac{M_{co,t_0}}{M_{o,t_0}}}_{p_{co}} \underbrace{\left(\frac{M_{ot} - M_{o,t_0}}{L_{c,t_0}}\right)}_{\gamma_{o,t-t_0}} \tag{3.11}$$

Where p_{co} are the shares of persons originating from o living in c at the period of reference and γ_{ot} are the immigrant-induced labour factor increase, from t-1 or from the reference period. In substance, the difference between these two expression is the definition of the shift that represent either a migration shock computed on a restricted period in the sample coverage, either the shock since the beginning of the study coverage. It implies that the

[1] See notably Caiumi and Peri (2024), Docquier et al. (2020), Jaeger et al. (2018) or Card (2001, 2009) for the use of various forms of the shift-share instrumental strategy in migration economics.

original eq. (3.11) captures the diaspora-induced cumulated migration flows since the period of reference t_0, while eq. (3.10) only captures recent migrations since $t-1$ but distributes it across units c according to the relative communities' presence in t_0. In other words, 2SLS estimates are more likely to capture the impact of the sum of previous migration shocks in the first case, and the recent shocks in the second. Summing the negative shocks and their recovery processes, the first is likely to detect a kind of weighted overall impact more than short-run influences, while the second is more likely to conflate workers' and capital mobility adjustments from recent migration shocks with short-run effects on native labor market outcomes, resulting in a positive bias: Migration patterns are indeed often stable (serially correlated) across the periods separated by relatively short intervals of time. It means that if the instrument predicts well migrations in a defined period, it also does for the preceding periods. It then captures both the influence of contemporary migration and the ongoing response of wages to previous migrations due to the dynamic nature of the effect (Jaeger et al., 2018). As a consequence, the estimate has to be interpreted as reflecting both the short-run and medium-run impacts of migrations, when the *original* version captures a longer history of migration shocks. Comparing their estimates is therefore instructive on the general dynamics of the migration impact, but they are also difficult to interpret. As mobility mechanisms may be subject to short-run rigidities in practice, mitigating the influence of adjustment from past shocks captured in the estimates helps to investigate the presence of such rigidities. A way to do so is to control for the influence of previous migrations instrumented by a form of lagged instrument, following Jaeger et al. (2018).

3.4.4 The levels of analysis where the shift-share strategy is risky

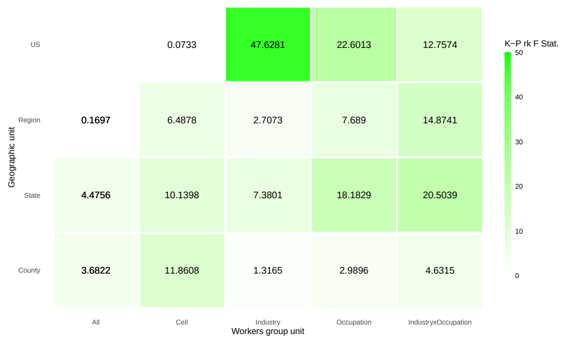
The shift-share instruments can be interpreted as functions that translate exogenous shocks to the desired level of analysis (Borusyak et al., 2025, p. 187). In the *many exogenous shifts* approach, the key assumption is the exogeneity of these shocks with respect to the outcome at the level of analysis. In the presence of workers' mobility in reaction to migration shocks, estimates at highly local levels capture the effects on the wages of native stayers. More aggregated levels include a net influence of mobility responses. To assess the net impact of migration on US wages, these last levels have therefore been preferred in the most recent literature: Due to the spatial and occupational sorting of natives, analysis at levels small in populations or geographically typically produces more local estimates that do not integrate broad complementarities. To draw more general conclusions, some exploit variations between broad regions to capture occupational and geographic mobility to a certain extent, while others adopt the national skill cell level to capture most of the geographic and occupational mobility of workers. This decision has implications for the shift-share instrumental approach regarding both the plausibility of its exogeneity and its predictive power of the migration flows. This section proposes a relatively short evaluation of the validity of these assumptions.

The number of shocks and the weights - The instrument embodies supposedly exogenous national shocks to the level of analysis considered (Borusyak et al., 2025, p. 187). These shocks represent the source of variations of the instrument. The more the shocks, the more the variations. A few random shocks may also be correlated with local wealth features by chance, while it is less likely to happen with many. The choice of the period of reference matters for the number of shocks used in the instrument. The section 3.B.4 table 3.B7 three last rows show the joint number of origins present between a reference period and three later periods. For example, there were 64 origins detected in 1950 that were also detected in 1990. It means that one can only measure shocks for those origins. Choosing 1950 as the reference period implies 64 to 65 shocks for each period. There was an increase in the diversity of migrations in the 70's, making the number of joint origins oscillate between 121 and 169 when taking the optimal 1980 reference period. I use this reference period for the main analysis. However, even if the number of shocks is the same for each level of analysis, the instrument applies different weights (share) to those shocks accordingly. The section 3.B.4 investigates the effective number of shocks used at each level for the *short-run* instrument (It measures approximately how many shocks actually contribute to the instrument, as some shocks may have very low diasporas' presence in the reference period), which gives an approximate idea of the variations in terms of origins behind the instrument. It tends to increase with spatial aggregation and skill disaggregation.

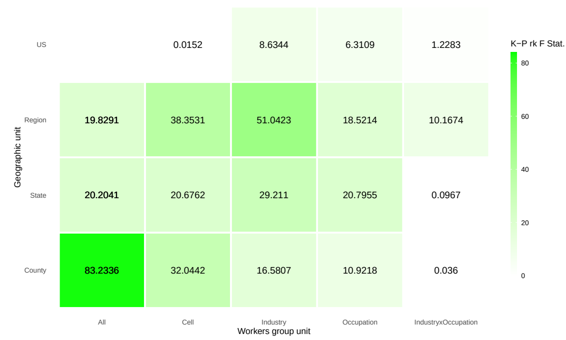
Strength and underidentification tests - Reliable 2SLS estimates suppose that the instrument has a strong explanatory power of the endogenous variable. The level of analysis interacted with the design of the instrument has a strong influence on its strength: The fig. 3.1 provides the first stage Kleibergen-Paap Wald rk F statistics[2] from 2SLS regressions using the two versions presented in section 3.4.3. As these matrices attest, the pattern of the usual (short-run) instrument form, presenting higher statistics in aggregated skill groups' levels, seems almost opposite to the *Cumulated* eq. (3.11) instrument equivalent statistics, which seems to perform better at aggregated geographic levels. One may believe that it is related to the different weights applied to the shocks[3], but decomposing the source of these divergences shows it is mostly the result of the shocks (section 3.B.4 fig. 3.B6). It suggests that recent immigrants are more sensitive to spatial rigidity and tend to stay in the same local area they came to in the short run, making the instrument a better predictor for recent migrations. Long-run migrations also capture the reallocation of earlier migrants to other places. On average, the panel form of the instrument is, therefore, a better predictor of migration exposure at a highly local level. However, the findings indicate that a shift-share approach relying only on skill groups' comparison to assess the impact of migrations is relatively risky for estimating short-run migration impacts.

[2]The Kleibergen and Paap (2006) F-statistics allow to perform underidentification tests where the null hypothesis states that the instrument has too weak power in explaining the endogenous variable to consider identification of the parameters.

[3]Indeed, the exogenous shock is the growth of origin o at the national level, g_{ot}, and not the origin o-induced local labor increase γ_{cot} that is mostly helpful in the interpretation of the instrument. As a consequence, the weight in the cumulated instrument is s_{cjo}, the share of migrants from origin o in the population of level cj at the period in 1980. The weights used in the more commonly used *short-run* version are actually the ratio between the population of the origin o at t_0 and the labor population at the level considered. Both instruments do not apply the same weight.



(a)



(b)

Note: Each squared cell represents the Kleibergen-Paap Wald rk F statistic (Kleibergen and Paap, 2006) derived from the first stage of the shift-share 2SLS regressions using the"cumulated" eq. (3.11) (a) or the "short run" eq. (3.10) version of the shift-share instrument (b).

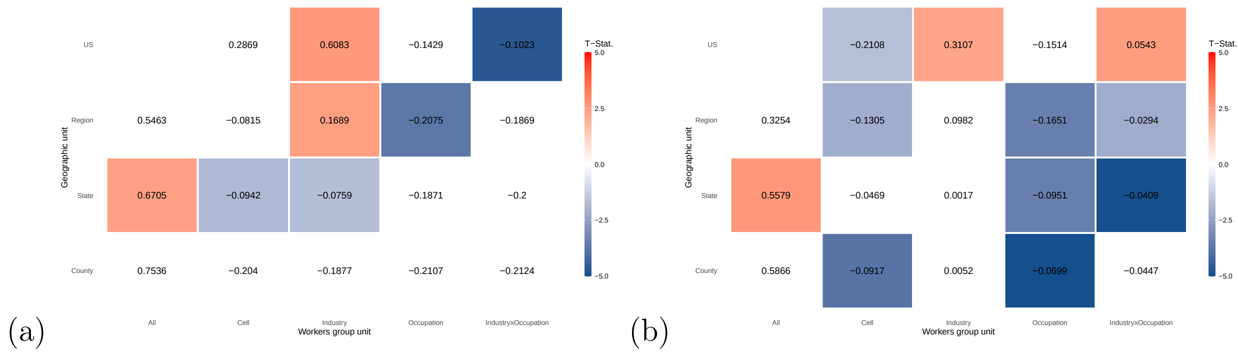
Figure 3.1: First stage Kleibergen-Paap rK F-Statistics, original (a) and modified (b) shift-share instruments

3.5 Regression analysis

3.5.1 Baseline OLS estimates of the migration short-run impact

The parameter δ_1 associated with the migrant share in eq. (3.12) is traditionally interpreted as a semi-elasticity of native wages to a migration-induced labor increase at the level of analysis. Its estimation relies on the comparison between exposed geographic units and non-exposed ones, hence the name of the spatial correlation method (Dustmann et al., 2016). The spatial correlation (and the skill-cell) approach achieves heterogeneous conclusions, while experimental studies exploiting sudden migration shocks appear to be more homogeneous regarding the short-run migration impacts (Borjas and Monras, 2016). The OLS coefficients at different levels of analysis are depicted in fig. 3.2 matrices for average and median wages. The left column represents the spatial correlation approaches, while the front row shows the results for a national skill cell approach, but without cell-time fixed effects. Other cells are mixed cases.

Aggregated levels are typically associated with positive or weak coefficients in (a), and highly disaggregated levels with negative, often insignificant, estimates. This difference is consistent with stronger competition between migrants and natives of similar skills and



Note: Each squared cell represents an OLS coefficient from a regression of the (log of) average or median hourly $1999 wage on the share of immigrant, both computed at different levels. The levels are defined geographically and according to workers' groups' definitions: All reflects the entire worker population in the area, Cell the experience education skill cells similar to Ottaviano and Peri (2012), Industry the detailed 1990 industries, Occupation the detailed 1990 occupations, and IndustryxOccupation all the joint components of these two last groups' definitions. Indices are computed at the state, County, and national (US) levels. Cells are colored according to the value of the *t-statistic*, with standard deviation clustered at the *cj* level. White cells represent insignificant estimates at a 10% threshold. Source: Author's elaboration on IPUMS 1980, 1990, 2000, 2007, and 2019 samples.

Figure 3.2: Migration exposures and average (a) and median (b) native wages, OLS estimates

complementarities at broader levels due to the diversity of skills[4]. As means may be sensitive to polarization and extreme values, matrix (b) imposes the median wage as an alternative: While positive estimates still characterize the pure skill cell and spatial approaches, mixed cases seem sensitive to this change, especially in experience-education groups.

Due to the circular causality between migration and wages induced by income pull-factors, the influence of migration on wages is likely to be overstated by the estimates. The next section applies the shift-share instrumental strategy in an attempt to reach identification.

3.5.2 Shift-share instrumental approach, the short-run impact of migration and insights from Compound estimates

This section replicates the shift-share instrumental strategy commonly used in the migration economics literature that consists here of instrumenting migration exposure by the predicted

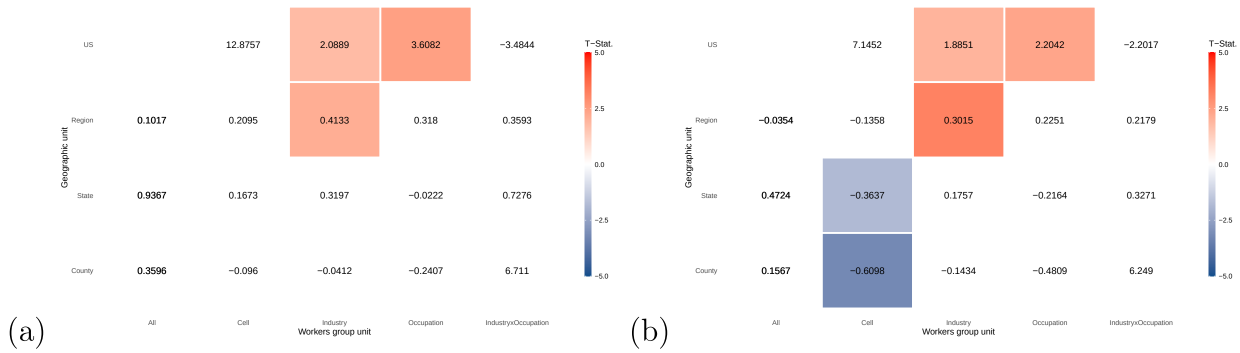
[4]this pattern is remarkably stable to resampling: when replicating the analysis with a 5-year interval between 2005 and 2019 (section 3.B.3 fig. 3.B3) or 15-year interval between 1990 and 2023 (fig. 3.5 (a)). Once not (directly) capturing changes in native populations in section 3.B.3 fig. 3.B3 (c), far more visible, suggesting that native mobility out of occupations and regions is a source of noise. It also shows that county-level estimates tend to be relatively similar in sign to more local geographic levels (metropolitan statistical areas and PUMAs). Note that the slight differences with the main results may originate from a less accurate measurement of hourly average wages for the years 2010 and 2015.

migrant share obtained while distributing each migration shock at the US level to regions and skill groups based on the diasporas' presence in 1980. I use the two versions of the instrument described in section 3.4.3: A "short-run" eq. (3.10) version that defines migration shocks as migration flows in the last ten years, and a "cumulative" eq. (3.2) instrument that is built on cumulative predicted migration flows since 1980. Both

Short-run instrument - In a first step, I apply the "short run" version of the instrument, capturing the migrant share predicted evolution during the last ten years. 2SLS estimates are displayed in fig. 3.3. Comparing them with OLS estimates shows a general increase in each category, at odds with income pull factors supposed to pollute the OLS estimation. Most of them are positive and insignificant, strongly echoing the general shift-share instrumental result of the spatial correlation literature (table 3.1), which I interpret as a sign of external validity. However, the internal validity of this instrumental strategy may suffer from migration flows serial correlation if there is a dynamic response of wages to immigration, as discussed in section 3.4.3: It would imply that the conditional immigration flows of the last ten years also capture the recovery process of the previous migration shock. For example, if migrations have an overall negative effect that progressively declines across time with the mobility of factors, these estimates are likely to be overestimated: As a complete recovery process can take more than 7 or 15 years (Edo, 2020; Jaeger et al., 2018; Cohen-Goldner and Paserman, 2011), it is likely that the short-run shift also captures the recovery from the shock of the previous period. As discussed

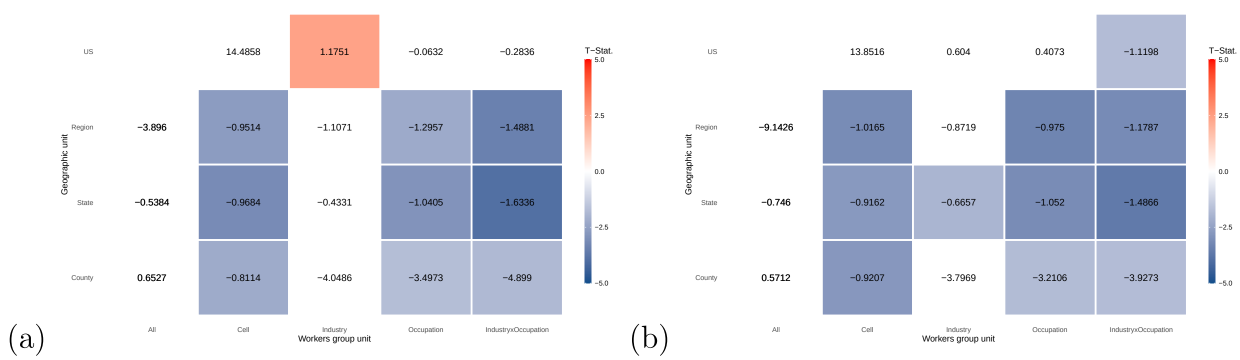
Cumulated instrument - The cumulated instrument predicts the total immigrant distribution since 1980 while the short-run instrument distributes the migrations of the last ten years. Knowing that the cycle of the effect of a specific migration shock is likely to last more than 10 years in the US context, the 2SLS estimates will provide a compound effect of the cumulated history of all migration shocks since 1980. In other words, they would reflect an average of all the responses of wages to current and past migration shocks still present in contemporary wages. While difficult to interpret and unable to identify the impact of migration shocks, they allow for testing the general direction of the effect and establishing if there is indeed a positive omitted variable bias in the previous estimates originating from including the recovery. The 2SLS main estimates are displayed in fig. 3.4 matrices. Both average (a) and median (b) native wages are negatively related to this cumulated predicted migration exposure, except at highly aggregated levels where the instrument has weaker predictive power. As discussed in section 3.4.3 and above, these 2SLS-estimated impacts of migration compound the influence of many successive migration shocks. In that sense, this specific spatial correlation method achieves the same general conclusions as structural and experimental studies, notably in the US (See table 3.3) with negative estimates: an overall negative impact of migration on native wages. In that sense, it reconciles the spatial correlation approach with experimental studies and defends a (positively biased) slight short-term negative effect.

National industry level - A persistent opposite pattern to this conclusion survives most of the specification: potential complementarities between foreign-born and native workers from the same national industry. An industry can reflect many different tasks, and this level is therefore more likely to be subject to occupational sorting in response to migration flows to benefit from comparative advantages. The section 3.4.4 also exhibits a better performance of the instrument for this specific level of analysis. The detected relationship between the national industrial migrant share and the median industrial wage is, however, close to 0. This element and the comparison of the other coefficients reveal that the subsequent average wage estimates are partially driven by a subsample of natives with high wages: most OLS coefficients become negative when focusing on the median native worker, and the positive effect at the industrial level is obliterated. These differences support that migrations benefit a subsample of workers, through general complementarities or a stronger geographic or occupational adaptability.



Note: Each squared cell represents a 2SLS coefficient from a regression of the (log of) median hourly \$1999 wage on the share of immigrant, both computed at different levels. The levels are defined geographically and according to workers' groups' definitions: All reflects the entire worker population in the area, Cell the experience education skill cells similar to Borjas (2003) or Ottaviano and Peri (2012), Industry the detailed 1990 industries, Occupation the detailed 1990 occupations, and IndustryxOccupation all the joint components of these two last groups' definitions. Indices are computed at the state, County, and national (US) levels. Cells are colored according to the value of the *t-statistic*, with standard deviation clustered at the cj level. White cells represent insignificant estimates at a 10% threshold. Shift-share instruments are defined according to eq. (3.10) (with units cj when required) using 1980 as the reference year. Source: Author's elaboration on IPUMS 1980, 1990, 2000, 2007, and 2019 samples.

Figure 3.3: Migration exposures and average (a) or median (b) native wages, short-run shift-share 2SLS estimates



Note: Each squared cell represents a 2SLS (b) coefficient from a regression of the (log of) average hourly \$1999 wage on the share of immigrant, computed at different levels. The levels are defined geographically and according to workers' groups' definitions: All reflects the entire worker population in the area, Cell the experience education skill cells similar to Ottaviano and Peri (2012), Industry the detailed 1990 industries, Occupation the detailed 1990 occupations, and IndustryxOccupation all the joint components of these two last groups' definitions. Indices are computed at the state, County, and national (US) levels. Cells are colored according to the value of the *t-statistic*. White cells represent insignificant estimates at a 10% threshold. Shift-share instruments are defined according to eq. (3.2) (with units cj when required) using 1980 as the reference year. Source: Author's elaboration on IPUMS 1980, 1990, 2000, 2007, and 2019 samples.

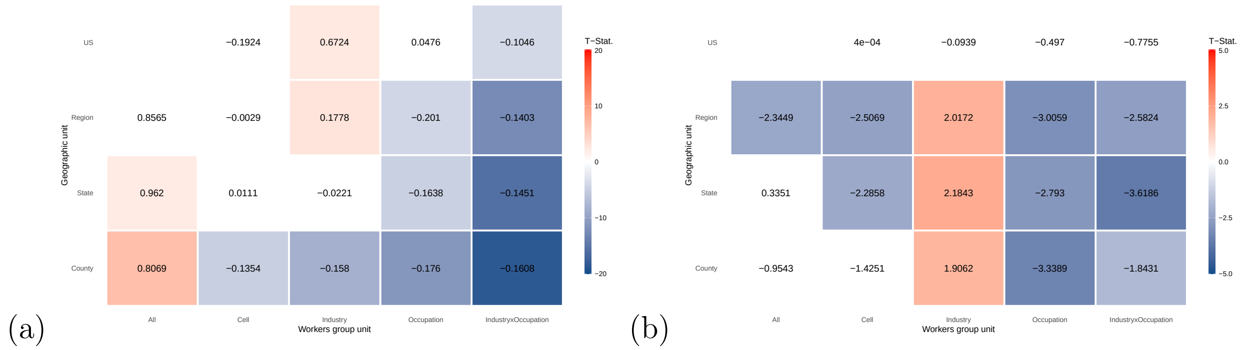
Figure 3.4: Cumulated migration exposures and average (a) or median (b) native wages, 2SLS estimates

Employment response to migrations - Interestingly, using the employment share as a dependent variable with long-run or median-run cumulated migration shocks in section 3.B.3 fig. 3.B4 (d) and fig. 3.B1 (b) depicts less or no impact on employment outcomes than short-run migration (section 3.B.3 fig. 3.B4 (b)). It suggests that the US labor market primarily adjusts to migration shocks through employment and then by wages.

3.5.3 Investigating the role of the definition of the shift

The contradiction between the results of the two instrument in the previous section supposes a relatively strong positive bias for its short-run version that is stronger than the bias in the cumulated version. Until now, the 10-year intervals of the data, that are often used in the literature, imposed a ten-year shift in the short-run version. Knowing that the dynamic should last more than 10 years (Jaeger et al., 2018; Edo, 2020), changing the definition of the short-run shift by using a longer timeline should then correct the estimates in the same direction as the cumulated instrument. Using comparable data and the most frequent availability of recent samples, this section replicates the initial instrumental strategy using intervals bigger than 15 years (1990, 2006, and 2023 samples). If previous estimates are indeed positively biased due to the recovery from the previous shock, adding 6 years should mostly detect a negative trend as the decrease preceding these recoveries should appear in the estimates. It allows to test if the main results of the cumulated instrument are related to older dynamics. The main results are displayed in fig. 3.5.

The OLS results are strongly similar to fig. 3.2, with more significant estimates, which supports the sample is comparable to the main sample used in this analysis. The 2SLS analysis in (b) indeed pictures a decrease of all estimates, consistent with this narrative, with a notable exception for the industrial level. Similar results are obtained with median wages (section 3.B.3 fig. 3.B1 (a)). Again, this result is consistent with the estimates capturing several negative shocks, whereas those in fig. 3.2 (b) capture more recent shocks, and potentially adjustment from early negative shocks. It also suggests a negative impact at the state aggregate level, which was not detected previously.



Note: Each squared cell represents an OLS (a) or shift-share 2SLS coefficient from a regression of the (log of) average hourly \$1999 wage on the share of immigrant, both computed at different levels. The levels are defined geographically and according to workers' groups' definitions: All reflects the entire worker population in the area, Cell the experience education skill cells similar to Ottaviano and Peri (2012), Industry the detailed 1990 industries, Occupation the detailed 1990 occupations, and IndustryxOccupation all the joint components of these two last groups' definitions. Indices are computed at the state, County, and national (US) levels. Cells are colored according to the value of the *t-statistic*. White cells represent insignificant estimates at a 10% threshold. Shift-share instruments are defined using 1980 as the reference year following eq. (3.10). Source: Author's elaboration on IPUMS 1980, 1990, 2006, and 2023 samples.

Figure 3.5: Migration exposures and (log of) average native wages, OLS (a) and short-run 2SLS (b) estimates - 1990, 2006, and 2023 samples

3.5.4 Accounting for past migrations and adjustment processes

The previous section exploits an alternative instrument to obtain information on the general sign of the effect of migration, but it is still uninformative on the magnitude of the elasticity of wages. In this section, I try to partially rehabilitate the "short-run" instrument by taking into account the recovery process of past migration shocks in the estimates, arising from serially correlated migration flows. To limit the influence of adjustment processes in the shift-share estimates of short-run migration impacts, which include capital and labour mobility in response to past migration flows, Jaeger et al. (2018) notably focus on periods prior to 1980 as they detect that migration flows became serially correlated later. As the table 3.B7 shows, it involves relying on a limited number of shifts (which increases the probability of endogeneity), and it is potentially less informative on more recent periods. Alternatively, a possible solution is to control for the instrumented lagged migrant shares, exploiting the low frequency of the data. Following the recommendation of Jaeger et al. (2018), I consider the following dynamic specification:

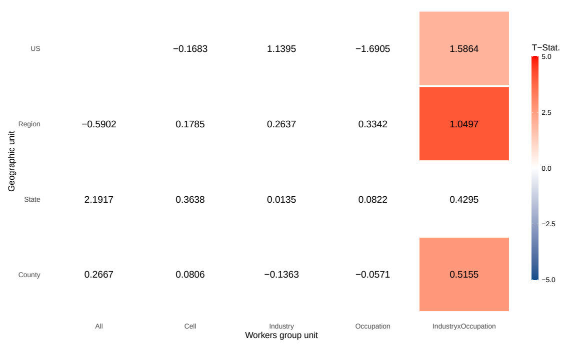
$$ln(w_{cjt}) = \beta_0 + \beta_1 \widehat{Migrants\ share}_{cjt} + \beta_2 \widehat{Migrants\ share}_{cj,t-1} + \gamma_j + \gamma_c + \gamma_t + \mu_{cjt} \quad (3.12)$$

Where $\widehat{Migrants\ share}$ terms represent contemporaneous and lagged shift-share predicted migration exposures. The instruments for migration exposure fitting this dynamic model are:

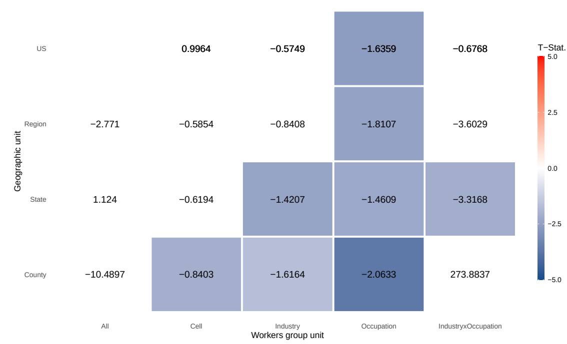
$$z'_{cjt} = \sum_{o=1}^{O} \frac{M_{cjo,t_0}}{M_{o,t_0}} \frac{M_{cjot} - M_{cjo,t-1}}{L_{cj,t-1}} \quad (3.13)$$

$$z'_{cj,t-1} = \sum_{o=1}^{O} \frac{M_{cjo,t_0}}{M_{o,t_0}} \frac{M_{cjot-1} - M_{cjo,t-2}}{L_{cj,t-2}} \quad (3.14)$$

Instrumenting the lagged share using the eq. (3.14) expression avoids endogeneity concerns arising from demand shocks. Due to the specific definition of the weights, I also control for the sum of shares for both periods (See section 3.B.4 for the definition of the shares in this context). It is, however, costly in terms of variations as it depends on the same reference periods as the contemporaneous migrant presence. Controlling for lagged migrant presence decreases the influence of the adjustment process from recent past migrations, but strong serial correlations may also produce inaccurate estimates for this reason: changes in migration flows are required to be able to detect a distinct influence from the two terms. Coefficients associated with the lagged instrumental share are systematically negative in fig. 3.6 (b), while the main coefficients of interest (a) are insignificant, except for industry-occupation groups. These results, relatively similar to the median worker estimates (fig. 3.7), strongly contrast with Jaeger et al. (2018), who found positive estimates for the lagged shares and negative coefficients for the contemporaneous independent variable. These authors argue that the lagged share captures the recovery process from previous migration shocks. The estimates here suggest that most of the negative influence is captured by the lagged share. While these results may also arise from serially strongly correlated migration flows, it can also reflect dominating adverse effects during the period covered in the present study, notably because of the appearance of rigidities: Shambaugh et al. (2017) mention a general decrease in both the geographic and the occupational mobility of workers in the last decades. These stylized facts induce a longer adjustment process that is consistent with this divergence. When inspecting the results for native employment shares in section 3.B.3 fig. 3.B4 (c), the estimates remain negative despite the lagged share controls. The combination of these results with previous estimations supports a short-run response of employment to migration shock, while wages react with a delay.



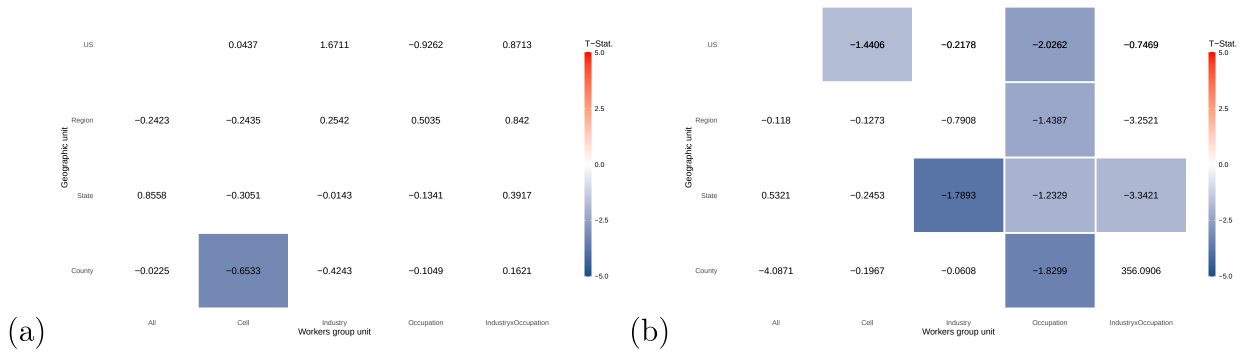
(a)



(b)

Note: Each squared cell represents a shift-share 2SLS coefficient from a regression of the (log of) average (a) and median (b) hourly \$1999 wage on the share of immigrant following the eq. (3.12) specification, at different levels of analysis. The levels are defined geographically and according to workers' groups' definitions: All reflects the entire worker population in the area, Cell the experience education skill cells similar to Ottaviano and Peri (2012), Industry the detailed 1990 industries, Occupation the detailed 1990 occupations, and IndustryxOccupation all the joint components of these two last groups' definitions. Indices are computed at the state, County, and national (US) levels. Cells are colored according to the value of the *t-statistic*. White cells represent insignificant estimates at a 10% threshold. Shift-share instruments are defined using 1980 as the reference year. Source: Author's elaboration on IPUMS 1980, 1990, 2000, 2007, and 2019 samples.

Figure 3.6: 2SLS estimates - Migration exposures and (log of) average (a) and median (b) native wages, controlling for lagged migration exposure



Note: Each squared cell represents a shift-share 2SLS coefficient from a regression of the (log of) median hourly $1999 wage on the share of immigrant (a) and the lagged share (b) at different levels of analysis. The levels are defined geographically and according to workers' groups' definitions: All reflects the entire worker population in the area, Cell the experience education skill cells similar to Ottaviano and Peri (2012), Industry the detailed 1990 industries, Occupation the detailed 1990 occupations, and IndustryxOccupation all the joint components of these two last groups' definitions. Indices are computed at the state, County, and national (US) levels. Cells are colored according to the value of the *t-statistic*. White cells represent insignificant estimates at a 10% threshold. Shift-share instruments are defined using 1980 as the reference year. Source: Author's elaboration on IPUMS 1980, 1990, 2000, 2007, and 2019 samples.

Figure 3.7: Shift-share 2SLS estimates - Migration and (log of) median native wages. Contemporaneous (a) and lagged (b) migration exposures.

3.6 Conclusion

The overall impact of migration depends on workers' complementarities, the migrants' contribution to innovation, their similarities in skill with other workers, and the speed of investment adjustment. According to the context and the composition of migrations, it can be an overall gain or loss of income for native workers; these may also be unequally distributed, as some workers in the destination country may be negatively or positively impacted according to their similarities with the migration inflows composition and how migration exposure is measured. This research investigated how this impact is deeply connected to the spatial and occupational sorting of natives in response to migrations, and how migration exposure is considered.

It estimated the impacts from both recent and long-run cumulated migration shocks on contemporary at different levels of geographic and skill aggregation to explore these mobility mechanisms: It allows for comparing effects when accounting for the two dimensions of native sorting. Moreover, as this mobility is deeply connected to the duration of the labor market adjustment to migration shocks, this research investigated the overall impact of migration through this dynamic, relying on different shift-share instrumental strategies. Doing so, it also investigated the performance of the shift-share approach in the period 1990-2019. It reveals a trade-off between aggregation and methodological concerns regarding the shift-share instrumental approach applied to the US context. More precisely, it highlights the poor performance of the shift-share strategy for skill-cell approaches at national levels of analysis and supports the choice of a right balance between the inclusion of complementarities arising from mobility and the good properties of the shift-share predictions. Intermediate levels exploiting both differences between skill and regional groups (mixed approaches) appear as safer alternatives.

Building on this, this research provides evidence of an overall negative impact, reconciling spatial correlation methods with recent findings exploiting sudden migration shocks. On the other hand, the findings also suggest positive effects at industrial levels, potentially arising from the existence of task complementarities between specific workers and competition effects on others, consistent with the idea of workers benefiting from Ricardian comparative advantages, notably through occupational mobility in the same industry. Turning to dynamic models, however, the results support a short-run adverse effect of migration on native employment share that progressively vanishes, while wages seem to be unaffected in the short run, at least for less mobile workers, and decrease in the medium-run. These differentiated effects are consistent with the presence of rigidities in the mobility of factors.

Further research is required to explore whether the heterogeneity of results in the literature that the current research was able to partially replicate with a single dataset is indeed explained by variations in the adjustment process of previous migration shocks, as the other results suggest. The study of the history of native mobility in the short run could provide an answer to this puzzle, as this heterogeneity may be rooted in the emergence of new rigidities in natives' sorting.

3.A Bibliography

Aleksynska, M., Peri, G., 2014. Isolating the network effect of immigrants on trade. The World Economy 37, 434–455.

Altonji, J.G., Card, D., 1991. The effects of immigration on the labor market outcomes of less-skilled natives, in: Abowd, J., Card, D. (Eds.), Immigration, Trade and the Labor Market. Routledge, p. 201–234.

Amior, M., Manning, A., 2018. The persistence of local joblessness. American Economic Review 108, 1942–1970.

Angrist, J.D., Kugler, A.D., 2003. Protective or counter-productive? labour market institutions and the effect of immigration on eu natives. The Economic Journal 113, F302–F331.

Aydemir, A., Borjas, G., 2007. Cross-country variation in the impact of international migration: Canada, mexico, and the united states. Journal of the European Economic Association 5, 663–708.

Basso, G., Peri, G., 2015. The Association Between Immigration and Labor Market Outcomes in the United States. IZA Discussion Paper 9436. IZA.

Beine, M., Docquier, F., Özden, Ç., 2011. Diasporas. Journal of Development Economics 95, 30–41.

Bonin, H., 2005. Wage and employment effects of immigration to Germany: Evidence from a skill group approach. IZA Discussion Papers No. 1875. IZA.

Borjas, G., Edo, A., 2021. Gender, Selection into Employment and the Wage Impact of Immigration. Technical Report.

Borjas, G.J., 2003. The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. The quarterly journal of economics 118, 1335–1374.

Borjas, G.J., 2014. Immigration economics. Harvard University Press.

Borjas, G.J., 2017. The wage impact of the marielitos: A reappraisal. Ilr Review 70, 1077–1110.

Borjas, G.J., 2018. The economics of immigration, in: The New immigrant in the American economy. Routledge, pp. 1–52.

Borjas, G.J., Grogger, J., Hanson, G.H., 2010. Immigration and the economic status of african-american men. Economica 77, 255–282.

Borjas, G.J., Monras, J., 2016. The labor market consequences of refugee supply shocks. Technical Report. National Bureau of Economic Research.

Borjas, G.J., Monras, J., 2017. The labour market consequences of refugee supply shocks. Economic Policy 32, 361–413.

Borusyak, K., Hull, P., Jaravel, X., 2022. Quasi-experimental shift-share research designs. The Review of Economic Studies 89, 181–213.

Borusyak, K., Hull, P., Jaravel, X., 2025. A practical guide to shift-share instruments. Journal of Economic Perspectives 39, 181–204.

Bosetti, V., Cattaneo, C., Verdolini, E., 2015. Migration of skilled workers and innovation: A european perspective. Journal of International Economics 96, 311–322.

Bratsberg, B., Raaum, O., 2012. Immigration and wages: Evidence from construction. The economic journal 122, 1177–1205.

Braun, S., Mahmoud, T.O., 2014. The employment effects of immigration: evidence from the mass arrival of German expellees in postwar Germany. The Journal of Economic History 74, 69–108.

Braun, S.T., Weber, H., 2021. How do regional labor markets adjust to immigration? a dynamic analysis for post-war Germany. Journal of International Economics 129, 103416.

Breunig, R., Deutscher, N., To, H.T., 2017. The relationship between immigration to australia and the labour market outcomes of australian-born workers. Economic Record 93, 255–276.

Caiumi, A., Peri, G., 2024. Immigration's Effect on US Wages and Employment Redux. Technical Report. National Bureau of Economic Research.

Card, D., 1990. The impact of the mariel boatlift on the miami labor market. Ilr Review 43, 245–257.

Card, D., 2001. Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. Journal of labor economics 19, 22–64.

Card, D., 2007. How immigration affects US cities. CReAM Discussion Paper Series 0711. Department of Economics, University College London.

Card, D., 2009. Immigration and inequality. American Economic Review 99, 1–21.

Carrasco, R., Jimeno, J.F., Ortega, A.C., 2008. The effect of immigration on the labor market performance of native-born workers: some evidence for spain. Journal of Population Economics 21, 627–648.

Carrington, W.J., De Lima, P.J., 1996. The impact of 1970s repatriates from africa on the portuguese labor market. Ilr Review 49, 330–347.

Cohen-Goldner, S., Paserman, M.D., 2011. The dynamic impact of immigration on natives' labor market outcomes: Evidence from israel. European Economic Review 55, 1027–1045.

Colas, M., 2019. Dynamic responses to immigration. Opportunity and Inclusive Growth Institute , 179–223.

David, H., Dorn, D., Hanson, G.H., 2013. The china syndrome: Local labor market effects of import competition in the united states. American Economic Review 103, 2121–68.

Docquier, F., Turati, R., Valette, J., Vasilakis, C., 2020. Birthplace diversity and economic growth: evidence from the us states in the post-world war ii period. Journal of Economic Geography 20, 321–354.

Dustmann, C., Fabbri, F., 2005. Immigrants in the british labour market. Fiscal Studies 26, 423–470.

Dustmann, C., Frattini, T., Preston, I.P., 2012. The effect of immigration along the distribution of wages. Review of Economic Studies 80, 145–173.

Dustmann, C., Preston, I., 2012. Comment: Estimating the effect of immigration on wages. Journal of the European Economic Association 10, 216–223.

Dustmann, C., Schönberg, U., Stuhler, J., 2016. The impact of immigration: Why do studies reach such different results? Journal of Economic Perspectives 30, 31–56.

Dustmann, C., Schönberg, U., Stuhler, J., 2017. Labor supply shocks, native wages, and the adjustment of local employment. The Quarterly Journal of Economics 132, 435–483.

d'Amuri, F., Ottaviano, G.I., Peri, G., 2010. The labor market impact of immigration in Western Germany in the 1990s. European Economic Review 54, 550–570.

D'Amuri, F., Peri, G., 2014. Immigration, jobs, and employment protection: evidence from europe before and during the great recession. Journal of the European Economic Association 12, 432–464.

Edo, A., 2016. How do rigid labor markets absorb immigration? evidence from france. IZA Journal of Migration 5, 7.

Edo, A., 2019. The impact of immigration on the labor market. Journal of Economic Surveys 33, 922–948.

Edo, A., 2020. The impact of immigration on wage dynamics: Evidence from the algerian independence war. Journal of the European Economic Association 18, 3210–3260.

Edo, A., Toubal, F., 2015. Selective immigration policies and wages inequality. Review of International Economics 23, 160–187.

Foged, M., Peri, G., 2016. Immigrants' effect on native workers: New analysis on longitudinal data. American Economic Journal: Applied Economics 8, 1–34.

Friedberg, R.M., 2001. The impact of mass migration on the israeli labor market. The Quarterly Journal of Economics 116, 1373–1408.

Glitz, A., 2012. The labor market impact of immigration: A quasi-experiment exploiting immigrant location rules in Germany. Journal of Labor Economics 30, 175–213.

Goldsmith-Pinkham, P., Sorkin, I., Swift, H., 2020. Bartik instruments: What, when, why, and how. American Economic Review 110, 2586–2624.

Hatton, T.J., 2015. United States immigration policy: The 1965 act and its consequences. The Scandinavian Journal of Economics 117, 347–368.

Howard, G., 2020. The migration accelerator: Labor mobility, housing, and demand. American Economic Journal: Macroeconomics 12, 147–179.

Hunt, J., 1992. The impact of the 1962 repatriates from algeria on the french labor market. Ilr Review 45, 556–572.

Hunt, J., Gauthier-Loiselle, M., 2010. How much does immigration boost innovation? American Economic Journal: Macroeconomics 2, 31–56.

Jaeger, D.A., Ruist, J., Stuhler, J., 2018. Shift-share instruments and the impact of immigration. Technical Report. National Bureau of Economic Research.

Kleibergen, F., Paap, R., 2006. Generalized reduced rank tests using the singular value decomposition. Journal of econometrics 133, 97–126.

Lewis, E., Peri, G., 2015. Immigration and the economy of cities and regions. Handbook of regional and urban economics 5, 625–685.

Llull, J., 2018. The effect of immigration on wages: exploiting exogenous variation at the national level. Journal of Human Resources 53, 608–662.

Mäkelä, E., 2017. The effect of mass influx on labor markets: Portuguese 1974 evidence revisited. European Economic Review 98, 240–263.

Manacorda, M., Manning, A., Wadsworth, J., 2012. The impact of immigration on the structure of wages: theory and evidence from Britain. Journal of the European economic association 10, 120–151.

McKenzie, D., Rapoport, H., 2010. Self-selection patterns in Mexico-US migration: the role of migration networks. the Review of Economics and Statistics 92, 811–821.

Mitaritonna, C., Orefice, G., Peri, G., 2017. Immigrants and firms' outcomes: Evidence from France. European Economic Review 96, 62–82.

Monras, J., 2020. Immigration and wage dynamics: Evidence from the Mexican peso crisis. Journal of Political Economy 128, 3017–3089.

Nedoncelle, C., Marchal, L., Aubry, A., Héricourt, J., 2025. Does Immigration Affect Native Wages? A Meta-Analysis. Working Papers 2025-07. CEPII. URL: https://www.cepii.fr/CEPII/en/publications/wp/abstract.asp?NoDoc=14651.

Ortega, J., Verdugo, G., 2014. The impact of immigration on the french labor market: Why so different? Labour Economics 29, 14–27.

Ortega, J., Verdugo, G., 2016. Moving up or down? Immigration and the selection of natives across occupations and locations. IZA Discussion Papers 10303. IZA.

Ortega, J., Verdugo, G., 2022. Who stays and who leaves? immigration and the selection of natives across locations. Journal of Economic Geography 22, 221–260.

Ottaviano, G.I., Peri, G., 2012. Rethinking the effect of immigration on wages. Journal of the European economic association 10, 152–197.

Peri, G., 2016. Immigrants, productivity, and labor markets. Journal of economic perspectives 30, 3–30.

Peri, G., Yasenov, V., 2019. The labor market effects of a refugee wave: Synthetic control method meets the mariel boatlift. Journal of Human Resources 54, 267–309.

Peters, M., 2022. Market size and spatial growth—evidence from Germany's post-war population expulsions. Econometrica 90, 2357–2396.

Pischke, J.S., Velling, J., 1997. Employment effects of immigration to germany: an analysis based on local labor markets. Review of Economics and Statistics 79, 594–604.

Polachek, S.W., Chiswick, C., Rapoport, H., 2006. The economics of immigration and social diversity. Emerald Group Publishing Limited.

Ruggles, S., Flood, S., Sobek, M., Backman, D., Cooper, G., Rivera Drew, J.A., Richards, S., Rodgers, R., Schroeder, J., Williams, K.C., 2025. IPUMS USA: Version 16.0. Minneapolis, MN: IPUMS 16. doi:https://doi.org/10.18128/D010.V16.0.

Sequeira, S., Nunn, N., Qian, N., 2020. Immigrants and the making of America. The Review of Economic Studies 87, 382–419.

Shambaugh, J., Nunn, R., Liu, P., Nantz, G., 2017. Thirteen facts about wage growth. The Hamilton Project 1.

Stark, O., Bloom, D.E., 1985. The new economics of labor migration. The american Economic review 75, 173–178.

Steinhardt, M.F., 2009. The wage impact of immigration in germany: New evidence for skill groups and occupations. The BE Journal of Economic Analysis Policy 11, 1–35.

Tumen, S., 2016. The economic impact of syrian refugees on host countries: Quasi-experimental evidence from turkey. American Economic Review 106, 456–460.

Vandenberghe, V., 2025. In Europe, arduous jobs fall on first-generation migrants: But later generations benefit from improved opportunities. De Economist , 1–24.

Winter-Ebmer, R., Zweimüller, J., 1996. Immigration and the earnings of young native workers. Oxford economic papers 48, 473–491.

Winter-Ebmer, R., Zweimüller, J., 1999. Do immigrants displace young native workers: the Austrian experience. Journal of Population Economics 12, 327–340.

Zorlu, A., 2016. Immigrants' occupational mobility—Down and back up again. IZA World of Labor. Institute for the Study of Labor (IZA).

Zorlu, A., Hartog, J., 2005. The effect of immigration on wages in three European countries. Journal of population economics 18, 113–151.

3.B Appendix

3.B.1 Samples

The main IPUMS (Ruggles et al., 2025) dataset used for the analysis is built using the 1980, 1990, and 2000 5% samples, as well as 2007 and 2019 ACS samples. These data were downloaded on November 27, 2025. Other results and statistics are computed using the 1950, 1960 and 1970 1% samples of the Census. Other Robustness checks are performed using either the 1980 and 1990 5% samples in combination with the 2006 and the 2023 ACS samples, or using the 2000 5% as well as the 2005, 2010, 2019, and 2023 ACS samples. These data were downloaded respectively on May 27, 2021, and on August 9, 2024.

3.B.2 Descriptive statistics

	Mean	SD	Min	Max	NxT
Regional level					
Av. hourly wage (natives)	17.74	2.19	14.03	23.16	45
Median wage (natives)	12.49	1.31	10.08	15.05	45
Migr. share	0.12	0.07	0.01	0.30	45
Empl. share (natives)	0.70	0.03	0.62	0.77	45
Population (thousands)	19,908.94	9,459.82	7,263.78	41,657.86	45
State level					
Hourly wage (natives)	17.15	2.81	11.55	29.26	255
Median wage (natives)	12.21	1.75	8.62	20.38	255
Migr. share	0.09	0.07	0.01	0.35	255
Empl. share (natives)	0.70	0.05	0.55	0.81	255
Population	3,513.34	3,990.91	275.80	25,793.86	255
County level					
Hourly wage (natives)	20.78	33.82	0.12	6,380.07	77,098
Median wage (natives)	16.08	11.47	0.11	1,575.00	77,098
Migr. share	0.16	0.17	0.00	0.99	77,496
Empl. share (natives)	0.72	0.20	0.00	1.00	77,496
Population	6.93	13.20	0.01	549.96	77,496
Cell level					
Hourly wage (natives)	22.63	12.77	7.20	142.34	232
Median wage (natives)	16.09	6.91	5.02	31.12	232
Migr. share	0.19	0.11	0.06	0.56	232
Empl. share (natives)	0.72	0.18	0.24	0.96	232
Population	3,235.22	2,448.55	11.99	10,582.20	232
Industry level					
Hourly wage (natives)	18.08	5.39	7.59	58.70	911
Median wage (natives)	13.11	3.95	4.61	25.70	911
Migr. share	0.14	0.08	0.01	0.53	915
Empl. share (natives)	0.82	0.09	0.00	0.99	915
Population	709.00	1,374.75	11.76	13,440.59	915
Occupation level					
Hourly wage (natives)	18.80	8.69	7.21	84.36	1,365
Median wage (natives)	14.51	6.37	4.48	57.43	1,365
Migr. share	0.14	0.09	0.01	0.59	1,365
Empl. share (natives)	0.82	0.09	0.34	0.98	1,365
Population	472.34	839.85	1.08	8,772.02	1,365
Industry-Occupation level					
Hourly wage (natives)	18.45	33.89	0.04	4,332.58	69,481
Median wage (natives)	15.04	13.15	0.02	1,445.40	69,481
Migr. share	0.19	0.16	0.00	0.98	70,072
Empl. share (natives)	0.81	0.19	0.00	1.00	70,072
Population	9.07	63.06	0.01	3,872.24	70,072

Note: The data are from the US IPUMS 1980, 1990, 2000, 2007, and 2019 samples described in section 3.B.1. SD is the standard deviation, and NxT is the number of observations. Descriptive statistics for Monetary variables are expressed in constant (1999) $. The employment share in the industry and/or occupation level is based on the most recent previous job.

Table 3.B1: Summary statistics - 1980-2019 samples, skill or geographic units

	Mean	SD	Min	Max	NxT
State Cell level					
Hourly wage (natives)	21.24	37.33	4.47	3,423.25	12,878
Median wage (natives)	15.65	11.57	1.13	1,043.20	12,878
Migr. share	0.13	0.13	0.00	0.83	12,881
Empl. share (natives)	0.72	0.18	0.00	1.00	12,881
Population	69.11	109.48	0.03	1,798.48	12,881
State industry level					
Hourly wage (natives)	17.10	13.47	0.29	1,306.10	43,442
Median wage (natives)	12.75	5.41	0.29	326.00	43,442
Migr. share	0.11	0.12	0.00	0.97	43,729
Empl. share (natives)	0.80	0.13	0.00	1.00	43,729
Population	17.29	47.24	0.02	1,619.74	43,729
State occupation level					
Hourly wage (natives)	17.90	13.81	0.37	1,202.16	59,760
Median wage (natives)	14.26	7.31	0.02	259.55	59,760
Migr. share	0.12	0.13	0.00	0.98	59,860
Empl. share (natives)	0.82	0.13	0.00	1.00	59,860
Population	12.47	30.66	0.01	1,092.49	59,860
State industry-occupation level					
Hourly wage (natives)	18.50	117.95	0.00	66,448.25	382,625
Median wage (natives)	15.89	119.60	0.00	66,448.25	382,625
Migr. share	0.24	0.20	0.00	1.00	389,640
Empl. share (natives)	0.81	0.23	0.00	1.00	389,640
Population	1.50	5.95	0.01	393.00	389,640

Note: The data are from the US IPUMS 1980, 1990, 2000, 2007, and 2019 samples described in section 3.B.1. SD is the standard deviation, and NxT is the number of observations. Descriptive statistics for Monetary variables are expressed in constant (1999) $. The employment share in the industry and/or occupation level is based on the most recent previous job.

Table 3.B2: Summary statistics - 1980-2019 samples, State skill units

	Mean	SD	Min	Max	NxT
County Cell level					
Hourly wage (natives)	20.78	33.82	0.12	6,380.07	77,098
Median wage (natives)	16.08	11.47	0.11	1,575.00	77,098
Migr. share	0.16	0.17	0.00	0.99	77,496
Empl. share (natives)	0.72	0.20	0.00	1.00	77,496
Population	6.93	13.20	0.01	549.96	77,496
County industry level					
Hourly wage (natives)	18.51	35.50	0.02	8,150.00	157,003
Median wage (natives)	14.34	26.62	0.02	8,150.00	157,003
Migr. share	0.19	0.17	0.00	0.99	159,409
Empl. share (natives)	0.80	0.18	0.00	1.00	159,409
Population	2.65	6.95	0.01	411.94	159,409
County occupation level					
Hourly wage (natives)	18.80	32.36	0.02	8,517.80	197,606
Median wage (natives)	15.44	22.24	0.01	8,030.00	197,606
Migr. share	0.20	0.19	0.00	0.99	199,875
Empl. share (natives)	0.81	0.19	0.00	1.00	199,875
Population	2.02	4.80	0.01	263.20	199,875
County industry-occupation level					
Hourly wage (natives)	19.48	50.82	0.00	16,300.00	452,470
Median wage (natives)	17.22	39.27	0.00	16,300.00	452,470
Migr. share	0.32	0.21	0.00	1.00	470,281
Empl. share (natives)	0.81	0.27	0.00	1.00	470,281
Population	0.54	1.50	0.01	116.37	470,281

Note: The data are from the US IPUMS 1980, 1990, 2006, and 2023 samples described in section 3.B.1. SD is the standard deviation, and NxT the number of observations. Descriptive statistics for Monetary variables are expressed in constant (1999) $. The employment share in the industry and/or occupation level is based on the most recent previous job.

Table 3.B3: Summary statistics - 1980-2019 samples, County skill units

	Mean	SD	Min	Max	NxT
Regional level					
Hourly wage (natives)	18.17	3.05	14.03	28.59	27
Median wage (natives)	12.45	1.39	10.08	15.03	27
Migr. share	0.14	0.08	0.02	0.30	27
Empl. share (natives)	0.71	0.03	0.65	0.78	27
Population (thousands)	21,079.38	10,004.48	8,614.64	42,843.07	27
State level					
Average wage (natives)	18.05	9.07	11.55	115.42	153
Median wage (natives)	12.18	1.90	8.62	22.25	153
Migr. share	0.10	0.08	0.01	0.35	153
Empl. share (natives)	0.72	0.04	0.58	0.81	153
Population (thousands)	3,719.89	4,256.63	283.71	25,267.51	153
County level					
Hourly wage (natives)	17.85	6.44	10.85	132.62	1,351
Median wage (natives)	12.61	2.41	7.75	26.30	1,351
Migr. share	0.12	0.10	0.00	0.59	1,351
Empl. share (natives)	0.72	0.05	0.50	0.87	1,351
Population (thousands)	421.28	646.57	53.60	6,564.78	1,351
Cell level					
Hourly wage (natives)	22.92	19.99	6.81	255.95	174
Median wage (natives)	15.90	6.76	4.96	31.37	174
Migr. share	0.19	0.11	0.06	0.59	174
Empl. share (natives)	0.72	0.18	0.26	0.96	174
Population	3,270.94	2,470.49	11.99	10,612.72	174
Industry level					
Hourly wage (natives)	18.26	9.73	7.59	151.15	683
Median wage (natives)	13.00	3.91	4.61	25.85	683
Migr. share	0.14	0.08	0.01	1.00	686
Empl. share (natives)	0.83	0.09	0.00	0.99	686
Population	719.13	1,420.33	11.65	13,482.30	686
Occupation level					
Hourly wage (natives)	18.75	13.32	7.19	338.08	1,028
Median wage (natives)	14.32	6.30	4.48	52.95	1,028
Migr. share	0.14	0.09	0.01	0.62	1,028
Empl. share (natives)	0.83	0.09	0.34	0.98	1,028
Population	477.09	871.18	1.08	10,274.11	1,028
Industry-Occupation level					
Hourly wage (natives)	18.27	40.67	0.04	5,471.70	47,426
Median wage (natives)	15.07	19.16	0.00	2,271.50	47,426
Migr. share	0.19	0.16	0.00	1.00	47,833
Empl. share (natives)	0.82	0.19	0.00	1.00	47,833
Population	10.07	67.35	0.01	3,766.87	47,833

Note: The data are from the US IPUMS 1990, 2006, and 2023 samples described in section 3.B.1. SD is the standard deviation, and NxT the number of observations. Descriptive statistics for Monetary variables are expressed in constant (1999) $. Differences in the numbers of observations are due to the absence of natives in specific skill groups. The employment share in the industry and/or occupation level is based on the most recent previous job.

Table 3.B4: Summary statistics - 1990-2023 samples, skill or geographic units

	Mean	SD	Min	Max	NxT
State Cell level					
Hourly wage (natives)	20.96	34.18	4.23	2,384.99	8,518
Median wage (natives)	15.50	8.05	2.28	240.92	8,518
Migr. share	0.14	0.13	0.00	0.85	8,519
Empl. share (natives)	0.73	0.18	0.00	1.00	8,519
Population	66.75	104.06	0.04	1,267.66	8,519
State industry level					
Hourly wage (natives)	17.71	59.61	1.31	7,323.62	24,742
Median wage (natives)	12.93	9.41	0.03	1,117.67	24,742
Migr. share	0.12	0.13	0.00	1.00	24,921
Empl. share (natives)	0.82	0.13	0.00	1.00	24,921
Population	19.43	52.89	0.02	1,586.10	24,921
State occupation level					
Hourly wage (natives)	17.97	32.39	0.13	3,120.28	33,779
Median wage (natives)	14.23	8.24	0.02	413.00	33,779
Migr. share	0.14	0.14	0.00	0.98	33,862
Empl. share (natives)	0.83	0.13	0.00	1.00	33,862
Population	14.08	33.89	0.01	1,190.12	33,862
State industry-occupation level					
Hourly wage (natives)	19.11	213.64	0.00	74,133.50	188,352
Median wage (natives)	16.66	243.72	0.00	74,133.50	188,352
Migr. share	0.26	0.21	0.00	1.00	192,420
Empl. share (natives)	0.82	0.24	0.00	1.00	192,420
Population	1.92	7.08	0.00	387.84	192,420

Note: The data are from the US IPUMS 1990, 2006, and 2023 samples described in section 3.B.1. SD is the standard deviation, and NxT the number of observations. Descriptive statistics for Monetary variables are expressed in constant (1999) \$. Differences in the numbers of observations are due to the absence of natives in specific skill groups. The employment share in the industry and/or occupation level is based on the most recent previous job.

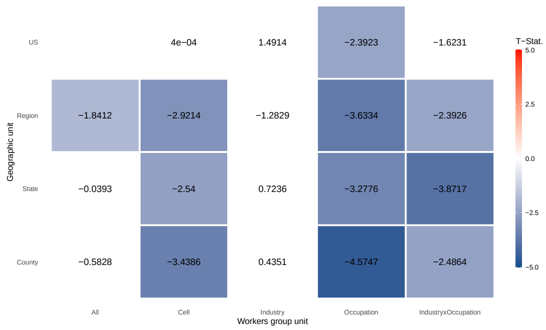
Table 3.B5: Summary statistics - 1990-2023 samples, State skill units

	Mean	SD	Min	Max	NxT
County Cell level					
Hourly wage (natives)	21.17	156.83	0.00	34,996.93	58,898
Median wage (natives)	16.14	55.18	0.00	11,623.75	58,898
Migr. share	0.17	0.18	0.00	1.00	59,277
Empl. share (natives)	0.73	0.21	0.00	1.00	59,277
Population	9.31	17.02	0.01	356.34	59,277
County industry level					
Hourly wage (natives)	18.46	69.13	0.01	17,902.35	107,821
Median wage (natives)	14.19	13.07	0.01	1,281.54	107,821
Migr. share	0.19	0.18	0.00	1.00	109,649
Empl. share (natives)	0.81	0.19	0.00	1.00	109,649
Population	4.01	10.39	0.01	449.74	109,649
County occupation level					
Hourly wage (natives)	18.50	49.63	0.02	12,034.92	136,108
Median wage (natives)	15.40	56.50	0.02	16,520.00	136,108
Migr. share	0.21	0.20	0.00	1.00	137,895
Empl. share (natives)	0.82	0.19	0.00	1.00	137,895
Population	3.05	6.86	0.01	280.66	137,895
County industry-occupation level					
Hourly wage (natives)	19.41	186.42	0.00	74,133.50	283,301
Median wage (natives)	17.21	145.53	0.00	74,133.50	283,301
Migr. share	0.32	0.22	0.00	1.00	294,627
Empl. share (natives)	0.81	0.27	0.00	1.00	294,627
Population	0.87	2.32	0.00	124.71	294,627

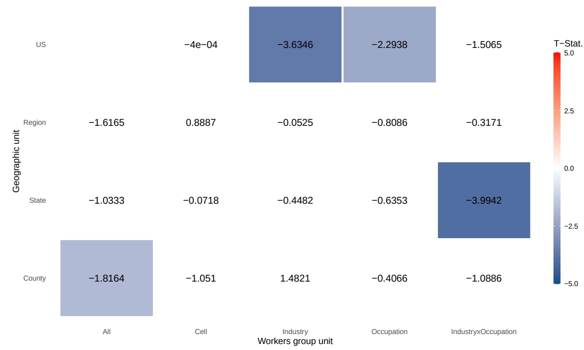
Note: The data are from the US IPUMS 1990, 2006, and 2023 samples described in section 3.B.1. SD is the standard deviation, and NxT the number of observations. Descriptive statistics for Monetary variables are expressed in constant (1999) \$. Differences in the numbers of observations are due to the absence of natives in specific skill groups. The employment share in the industry and/or occupation level is based on the most recent previous job.

Table 3.B6: Summary statistics - 1990-2023 samples, County skill units

3.B.3 Analysis appendix



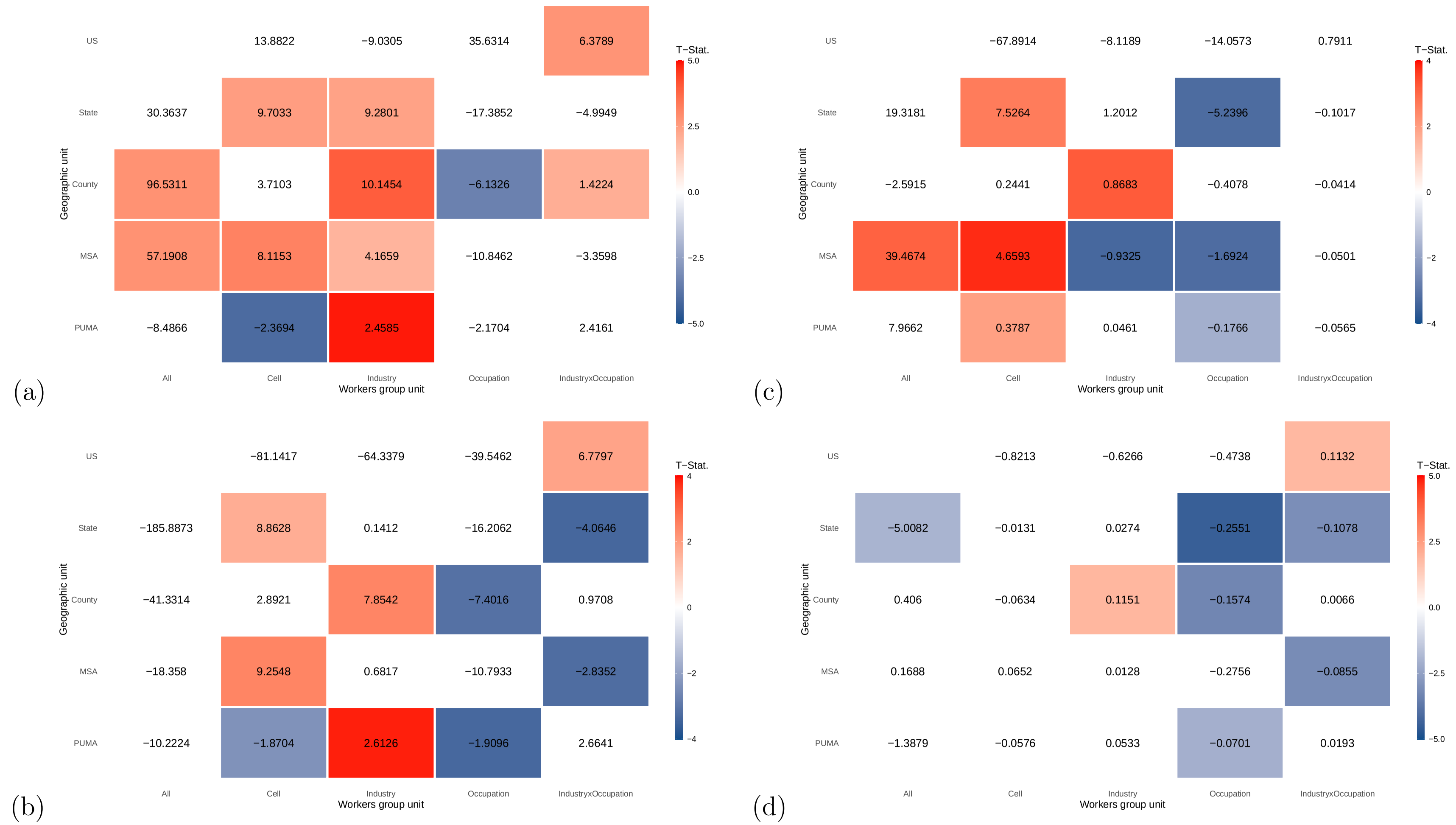
(a)



(b)

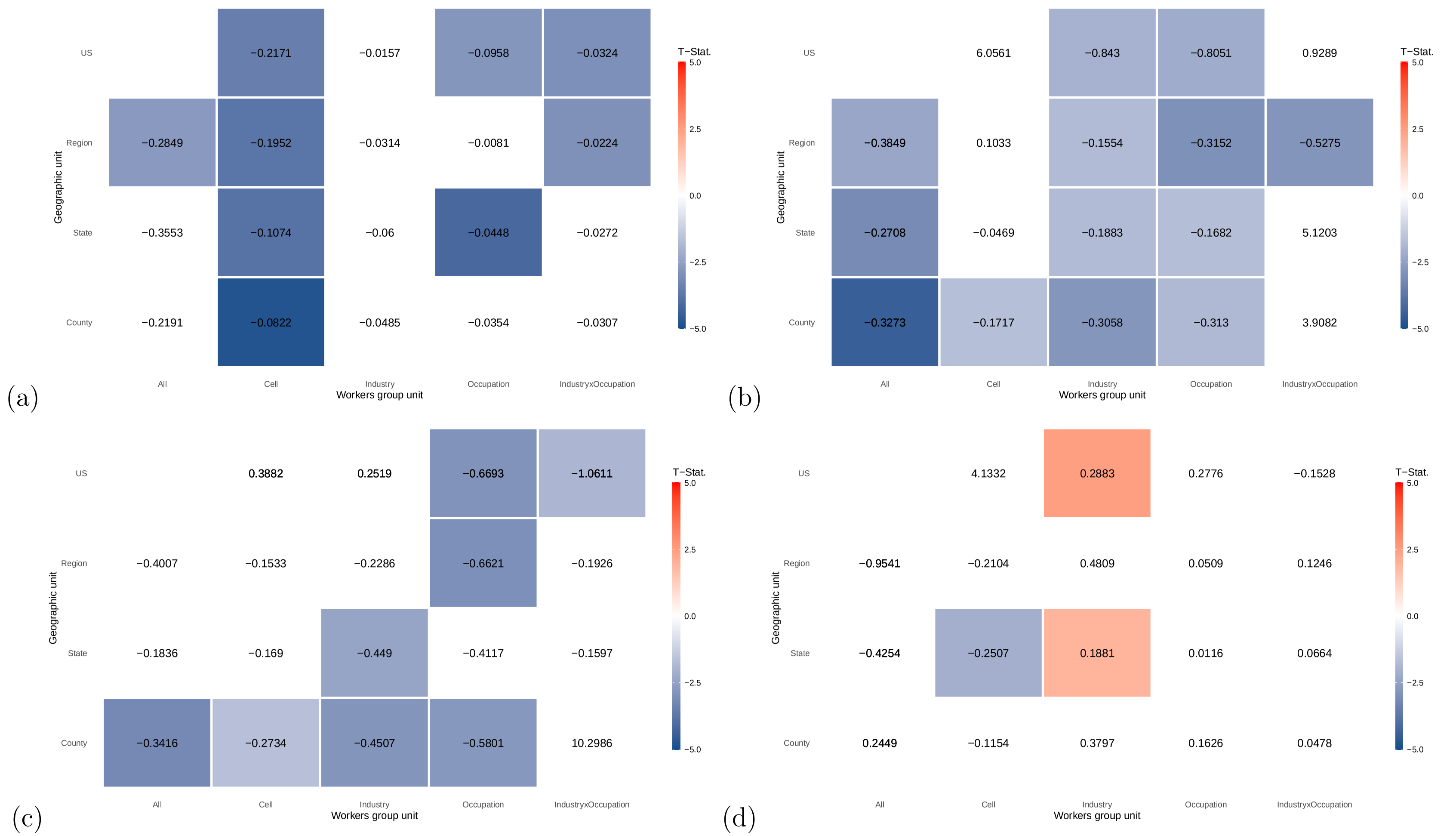
Note: Each squared cell represents a shift-share 2SLS coefficient from a regression of the (log of) average (a) and median (b) hourly \$1999 wage on the share of immigrant, both computed at different levels. The levels are defined geographically and according to workers' groups' definitions: All reflects the entire worker population in the area, Cell the experience education skill cells similar to Ottaviano and Peri (2012), Industry the detailed 1990 industries, Occupation the detailed 1990 occupations, and IndustryxOccupation all the joint components of these two last groups' definitions. Indices are computed at the state, County, and national (US) levels. Cells are colored according to the value of the *t-statistic*. White cells represent insignificant estimates at a 10% threshold. Shift-share instruments are defined using eq. (3.10) and 1980 as the reference year. Source: Author's elaboration on IPUMS 1990, 2006, and 2023 samples.

Figure 3.B1: Migration exposure, median native wage (a) and employment share (b), 2SLS estimates - 1990, 2006 and 2023 samples



Note: Each squared cell represents an OLS (a and b) or shift-share 2SLS (c) coefficient, multiplied by 100, from a regression of the (log of) average hourly $1999 wage on the ratio of immigrant and native working populations. Specifications (a) and (c) follow the eq. (3.12) baseline; specification (b) fixes the native population at its 2000 value. The levels are defined geographically and according to workers' groups' definitions: "All" reflects the entire worker population in the area, Cell the Ottaviano and Peri (2012)'s experience-education skill cells, "Industry" the detailed 1990 industries, "Occupation" the detailed 1990 occupations and "IndustryxOccupation" all the joint components of these two last groups. As areas, "US" is associated with the national level results, "State" the (US) states level, "County" the county level, "MSA" the Metropolitan statistical area level and "PUMA" the Public Use Microdata Area level, that is the smallest geographic unit available. The estimates are weighted according to the population inside these groups to retrieve results representative of the US. Darker cells represent higher *t-statistics*. White cells are insignificant at 10%. Source: Author's elaboration on IPUMS 2000-2019 ACS data.

Figure 3.B3: Estimated coefficients matrices - 2005-2019 (a) and 2015-2019 (b) samples baselines estimates, 2005-2019 estimates with fixed native population (c), and 2015-2019 2SLS estimates (d)



Note: Each squared cell represents an OLS (a) or shift-share 2SLS coefficient from a regression of the native employment rate on the share of immigrants, both computed at different levels. Instruments based on eq. (3.10) The levels are defined geographically and according to workers' groups' definitions: All reflects the entire worker population in the area, Cell the experience education skill cells similar to Ottaviano and Peri (2012), Industry the detailed 1990 industries, Occupation the detailed 1990 occupations, and IndustryxOccupation all the joint components of these two last groups' definitions. Indices are computed at the state, County, and national (US) levels. Cells are colored according to the value of the *t-statistic*. White cells represent insignificant estimates at a 10% threshold. Shift-share instruments are defined using 1980 as the reference year. Employment rates at occupational and industrial levels are computed according to the population working in the sector and/or the occupation during the previous year. Source: Author's elaboration on IPUMS 1980, 1990, 2000, 2007 and 2019 samples.

Figure 3.B4: Migration exposures and native employment rate, OLS (a), panel without (b) and with lagged share control (c) 2SLS estimates, and original shift-share 2SLS estimates (d)

3.B.4 Shift-share appendix

Effective number of shock

Effective number of shocks and summary statistics -Exogeneity requires numerous shocks: Even if these are fully random, few shocks may be correlated by chance with unobservable features. The shocks are also more likely to be exogenous if the shift level and the level of analysis (the same as the shares) are distant from each other. National-level shocks are the most distant feasible shifts and are therefore usually preferred in the migration literature using a specific country as a case study. However, they still imply that the level of the analysis and the shares are sufficiently distant, such that national shocks cannot affect wages at that level other than through diaspora-induced movements. This exogeneity assumption is likely to be violated if the skill cells or the regions used by the researchers aren't numerous. For this reason, Borusyak et al. (2022) suggest a systematic review of the moments of the shocks and their respective contribution to the instrument: Even with numerous origins, aggregated levels may actually still induce unreliable 2SLS estimates if migrations are not sufficiently diversified across regions and categories, as specific birthplaces would contribute more to the instrument. Consider the effective number of shifts (i.e., their number weighted by the average share of each origin at the level of analysis):

$$Eff.\ number\ of\ shocks = \frac{N_{cj}^2}{\sum_o (\sum_{cj} s_{cjo,t_0})^2} \tag{3.B1}$$

This expression is the equivalent of the inverse Herfindahl index of the average weight of each origin o across all observations, $S_o = \frac{1}{N_{cj}} \sum_{cj} s_{cjo,t_0}$. Borusyak et al. (2022) use this effective number to evaluate the industrial diversity of the US and picture the variations at the core of the instrumental strategy of David et al. (2013). Computing this effective number of shocks for different levels and periods of reference allows for comparison: The higher the Herfindahl index, the more concentrated the migration flows at the level of analysis. Its inverse, therefore, provides a measure of historical diaspora diversity. As each origin flow participates proportionally to its historical share, it allows for inspecting if a subsample of birthplaces drives the main results. The table 3.B7 provides summary statistics (mean, standard deviation, and interquartile range) of the shifts g_{ot} at the national skill cell level, weighted by the average shares. The decrease in the shocks means and variations across periods of reference t_0 highlights the increasing US migration rates across time. The last rows show the number of communities present both at the reference period and during the sample year. The sudden increase between 1970 and 1980 reflects two specific shocks: The Immigration and Nationality Act, resulting in increasing extra-European migration flows (Hatton, 2015), and the emergence of new countries, notably due to the decolonial process. This phenomenon implies a higher number of shocks when using more recent years as reference periods. It is worth noting that these values are specific to the period of reference and therefore have implications for any research at any level relying heavily on a shift-share strategy[5]. Unsurprisingly, it results in lower numbers of effective shocks for anterior periods. Additionally, these changes in the number of origins are temporally concomitant with the apparition of serially correlated continuous migrations to the US described by Jaeger et al. (2018). The use of a shift-share instrument combined with the absence of serially correlated migrations is supposed to ensure that estimates do not capture adjustement process from previous migration waves. In the US context, these findings indicate that this identification strategy requires relying on a lower number of shocks.

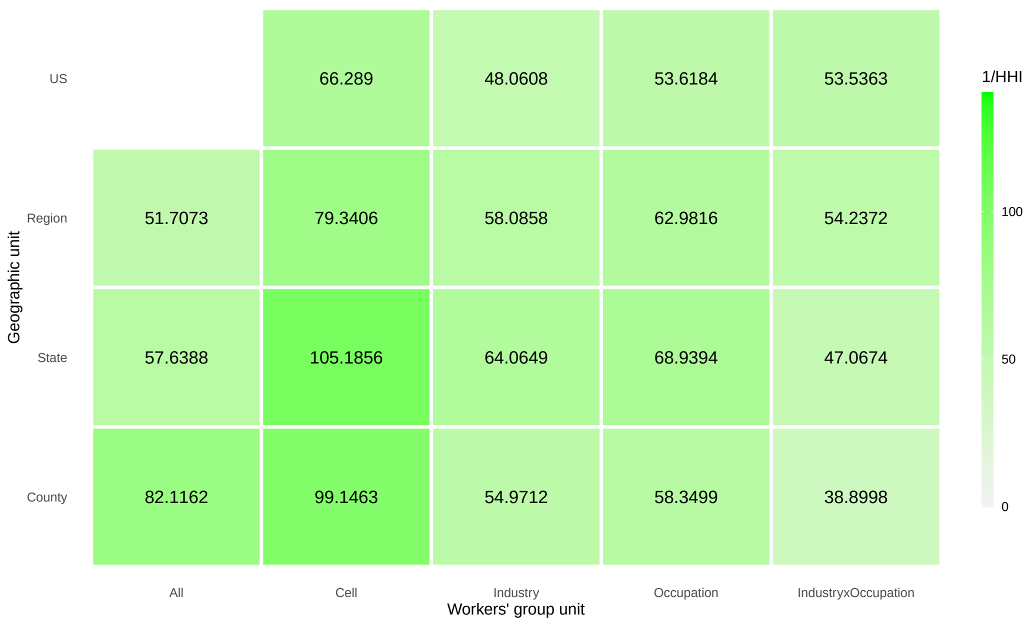
[5]These values are specific to the US, but the phenomenon is likely to be at core in many other contexts.

	Reference year t_0					
Statistic	1950	1960	1970	1980	1990	2000
mean	5.3267	4.4147	3.3636	1.8160	0.7555	0.1206
s.d.	33.5663	19.2786	9.0025	4.1724	1.6970	0.6004
Interquartile range$_{75\%-25\%}$	0.9336	0.9147	2.6422	2.7190	1.2140	0.5664
Eff. # of shocks	39.0674	46.3855	53.2665	66.2890	72.5128	56.3295
# time x shocks	193	196	208	422	238	246
# birthplaces, 1990	64	65	69	169	-	-
# birthplaces, 2006	65	66	70	121	114	122
# birthplaces, 2023	64	65	69	132	124	124

Note: This table provides summary statistics (mean, standard deviation, and interquartile range) of the shocks defined in eq. (3.2) for different periods of reference at the national skill-cell level. The Effective # of shocks is computed using eq. (3.B1) formula. # time x shocks is the total number of shocks. # birthplaces provides the common number of immigrant origins between the period t_0 and the period t considered. Source: Author's computation on 1950, 1960, 1970, 1980, 1990, 2000, 2006, and 2023 IPUMS data.

Table 3.B7: Shifts summary statistics at skill cell level

The effective number of shocks increases after 1970 and peaks in 1990. They remain more or less three times lower than the values obtained by Borusyak et al. (2022) in their tests on the instruments of David et al. (2013). In this research, the chosen reference year is 1980, which provides relatively numerous effective shocks and is relatively distant from the periods of the main sample. A visual inspection of the Herfindahl index computed from other levels of analysis for this period in fig. 3.B5 supports a better performance at spatially local levels. This heterogeneity across levels of analysis may sound surprising. A classic strategy to improve the exogeneity of the shift-share instrument in migration studies is to exclude the most represented community from its computation. While it makes the estimates more local, it increases the number of effective shocks with an unequal distribution of migrants. The section 3.B.3 fig. 3.B7 shows it indeed increases the sources of variations at highly aggregated levels, but not that much at the State-cell level, which is the optimum in both heatplots. This difference suggests that spatially disaggregated levels decrease the weight of the main migrant communities because they are spatially strongly unequally distributed. It means that the shift-share instrument is more likely to be valid at spatially disaggregated skill-cell levels, but it also implies that this level is less likely to reflect the impact of representative migrations to the US.



Note: Each squared cell represents the effective number of shocks (inverse Herfindahl index) derived from eq. (3.B1) at a specific level of analysis cj. Shares are computed from the migrant populations in 1980. Industries and occupations are defined according to their 1990 definition. Source: Author's computation on 1980, 1990, 2006, and 2023 IPUMS data.

Figure 3.B5: Number of effective shocks - Inverse Herfindahl indices of the 1980 average birthplace shares, by level of analysis

Discussion on the two shift-share varieties

The main difference between the *original* eq. (3.2) shift-share instrument and its panel-adapted version, rewritten as the change in predicted migration flows divided by past population in eq. (3.B2), is that only immigrants' inflows between t and $t-1$ are distributed across groups/regions in the second case (and not the whole inflows since the reference period).

$$\frac{\Delta\widehat{M_{ct}}}{L_{c,t-1}} = \sum_{o=1}^{O} \frac{M_{co,t_0}}{M_{o,t_0}} \frac{M_{cot} - M_{co,t-1}}{L_{c,t-1}} \tag{3.B2}$$

It decreases the influence of the cumulated migration shocks in the estimates. However, another major difference lies in what seems to be the shifts and the shares. The instrument is still a weighted average of the national inflow rates from each origin o, but it is here defined as an immigrant-induced increase in the workforce between the two periods. The appearance of the formula gives the impression that the weight of each shock is equal to the others by construction. However, the true share is actually difficult to interpret, as the immigrant population does not solely determine the shock. In a sense, it is equivalent to applying the weight $\frac{M_{co,t_0}}{L_{c,t-1}}$ to an immigrant-defined shock. As a consequence, the weights are not solely defined by the past diaspora presence in city j, but also by the population size at time $t-1$. It means that the missing share is not (only) the native share, but a very specific concept: the change toward the new representative migrant from the reference period to $t-1$. While the missing share approach could eventually be avoided with the original instrument by redefining the origins' shares according to the immigrant population, the difference in times would imply an omitted variable bias if one does not control for the sum of the shares as recommended by Borusyak et al. (2022).

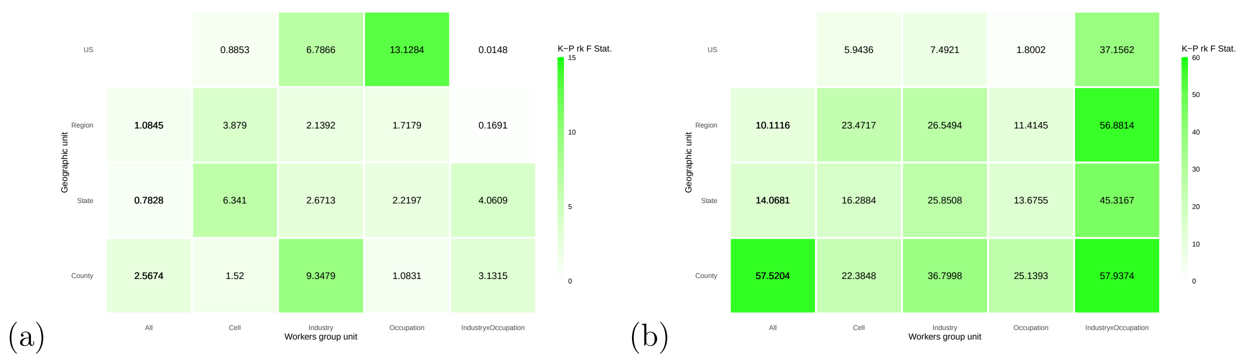
Underidentification tests

This difference induces different values and interpretations, but also properties, having important consequences for the analysis: The fig. 3.B5 shows that either of the two instruments can dominate the other in strength according to the level of analysis, but also that one or the other, at least for the period considered in the present study, can seriously damage the ability to identify the parameter δ_1. To ensure that this divergence is driven by the timeline of the migration shocks and not the particular definition of shares in eq. (3.10), I compute two mixed versions of the shift-share instruments:

$$z_{ct}^{a} = \sum_{o=1}^{O} \frac{M_{co,t_0}}{M_{o,t_0}} \left(\frac{M_{ot} - M_{o,t_0}}{L_{c,t-1}} \right) \tag{3.B3}$$

$$z_{ct}^{b} = \sum_{o=1}^{O} \frac{M_{co,t_0}}{M_{o,t_0}} \left(\frac{M_{ot} - M_{o,t-1}}{L_{c,t_0}} \right) \tag{3.B4}$$

Replicating the 2SLS estimates with these mixed instruments and extracting their respective first-stage K-P rk F Statistics in fig. 3.B6 allows for inspecting which decomposed change in the formula explains the differences in the fig. 3.1 patterns. As matrix (b) shows, the instrument's stronger performance is entirely driven by restricting it to the distribution of recent immigrants and not the size of the population in the previous period, which results in a general decrease in the statistics (a), especially at local levels.



Note: Each squared cell represents the Kleibergen-Paap Wald rk F statistic (Kleibergen and Paap, 2006) derived from the first stage of the shift-share 2SLS replications of fig. 3.4 (b) regressions, using eq. (3.B3) (a) and eq. (3.B4) (b) modified shift-share instruments.

Figure 3.B6: First stage Kleibergen-Paap rK F-Statistics, shift's denominator (a) and numerator (b) modified instruments

Note: Each squared cell represents the effective number of shocks (inverse Herfindahl index) from eq. (3.B1) at a specific level of analysis cj. Shares are computed from the migrant populations in 1980. Industries and occupations are defined according to their 1990 definition. Source: Author's computation on 1980 (for values), 1990, 2006, and 2023 (for included birthplaces) IPUMS data.

Figure 3.B7: Inverse Herfindahl index of the 1980 average birthplace shares excluding Mexicans, by level of analysis

Conclusion

Globalization, in its historical and contemporary forms, has shaped economic outcomes and human interactions across regions and generations. The three chapters of this thesis collectively illustrate the influence of specific aspects of globalization related to political and cultural integration, as well as the mobility of people and goods. The first chapter demonstrates that early political integration during the Roman era left persistent legacies in trade and economic connectivity across Western Europe, highlighting how historical globalization shocks can shape the long-run patterns of interaction between human communities. The second chapter argues that, in modern contexts, the coexistence of culturally and economically diverse migrant groups within occupations and sectors generates localized productivity gains, emphasizing the importance of complementarities and interaction effects. Finally, the third chapter examines the short- and medium-run responses of native labor markets to migration shocks, revealing the sensitivity of the estimated migratory impact on wages to methodological considerations.

While seemingly unrelated, these findings highlight that the economic effects of globalization are deeply associated with historical and institutional contexts, the structure of local interactions, and the capacity of individuals and communities to adjust to new flows of people and ideas. By integrating evidence from past and present episodes, this thesis contributes to a more nuanced understanding of globalization as a dynamic, multi-layered process, whose consequences cannot be fully understood without accounting for historical legacies and adjustment mechanisms. In doing so, it highlights the enduring importance of migration, cultural exchange, and political integration as fundamental channels through which globalization shapes economic ties and outcomes.

Funding

The views expressed herein are exclusively those of the author.