PhD-FDEF-2025-030
The Faculty of Law, Economics and Finance

DISSERTATION

Defence held on 12/12/2025 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN SCIENCES ECONOMIQUES

by

Ariane GORDAN
Born on 25 July 1991 in Nuremberg (Germany)

ESSAYS ON THE ECONOMICS OF MIGRATION AND MARRIAGE

Dissertation defence committee

Prof. Dr Arnaud Dupuy, Dissertation, supervisor
Professor, Université du Luxembourg

Prof. Dr Konstantinos Tatsiramos, Chairman
Professor, Université du Luxembourg

Dr Cesare Riillo
Senior Researcher, STATEC

Asst. Prof. Dr Shahdi Farahzadi
Assistant Professor, University of Liverpool Management School

Prof. Dr Joan Llull
Research Professor, Institute of Economic Analysis (IAE-CSIC)

Acknowledgements

I am deeply grateful to many people without whom I would not have been able to undertake this endeavor. First and foremost, I would like to thank my thesis supervision committee: Konstantinos Tatsiramos, Cesare Riillo, and, in particular, my principal supervisor Arnaud Dupuy; for their trust, patience, and thoughtful guidance during this intense journey. Throughout the last four years, Arnaud has generously shared his expertise and advice while constantly challenging me to grow as an independent researcher. It may not have always seemed that way, but I enjoyed this process immensely, despite any "growing pains"!

My sincerest gratitude further goes to the two external members of my defense jury, Shadi Farahzadi and Joan Llull, for agreeing to share their expertise and perspective on my thesis. Joan was also kind enough to host me at UAB last fall, which turned out to be a decisive moment in my PhD in many ways. Thank you, Joan, for warmly welcoming me to your research group, and for the insightful exchanges and thoughtful advice. I also want to express my heartfelt thanks to the PhD community at UAB, particularly to Carlos, Sergi, Sophie, and Nada, for adopting me into their IDEA family for a while.

From Luxembourg, I would like to thank my co-authors Frederic Doquier and Michel Tenikue for creating a collaborative atmosphere in which I always felt free to explore and contribute equally. I learned a lot from working beside you, not least the importance of openness and mutual respect in research. I am also deeply thankful to the DEM faculty, including Vincent Anesi, Michel Beine, Pierre Picard, and Skerdi Zanaj, for their valuable feedback on my work and genuine advice on numerous occasions. Further, I would like to acknowledge the administrative support by the DEM/ACROSS/DSEFM team, and especially Noemie Curtois, who we could always count on for organizational and emotional support. I gratefully acknowledge the financial support by the Luxembourg National Research Fund under a PRIDE grant for the ACROSS Doctoral Training Unit.

I feel extremely lucky to have met so many kind and talented people during my time at Uni.Lu, who made the good times so much more enjoyable and helped me overcome the difficult times. To Aleksa, Aurelie, and Dara, I am grateful to have had such amazing office mates, and to get to keep such precious friendships; our countless discussions about our research are reflected copiously in this thesis. A special thank you also goes to my dear friend and co-author Camila, who is casually one of the kindest people I know. Your unconditional support and encouragement have motivated me beyond measure. I want to further thank my colleagues and friends at the department and beyond, including Adrij, Andrei, Anass, Cecilia, Cristina, Elisabeth, Emma, Etienne, Evie, Lauriane, Maria, Matteo, Mireia, Raian, Sandra, Sarai, Wajdi, and many more, for making this journey a whole lot more fun.

Finally, I am beyond grateful for the support of my family and long-time friends. It is impossible to express the extent of my appreciation for the unconditional love and support I have received from my parents, Ali and Claudia. *Mama, Papa: Eure liebevolle und bedingungslose Unterstützung gibt mir eine unbeschreibliche Sicherheit und Vertrauen in meinen Weg. Dafür bin ich euch unermässlich dankbar.* I want to further thank my sister Sina, who was the first in our family to embark on an academic journey and provided loving advice and perspective whenever I needed it. *Danke, Schwesterherz.* Lastly, to my dear friends "back home": *Conny, Lisa, Melli, Silvia, Stephi – Danke für diese besondere Freundschaft, die keine Distanzen kennt.*

Contents

Abstract **11**

Declaration of Co-Authorship **13**

Declaration of Use of Generative AI **14**

1 Marriage Matters: Internal Migration and Marital Sorting in Indonesia **15**
1.1 Introduction . . . 16
1.2 Data . . . 20
1.2.1 Sample . . . 20
1.2.2 Descriptive evidence . . . 21
1.3 Model . . . 24
1.3.1 Two-sided maximization problem . . . 25
1.3.2 Equilibrium . . . 27
1.3.3 Identification . . . 28
1.3.4 Spousal and location types . . . 28
1.4 Estimation . . . 29
1.5 Estimation results . . . 30
1.5.1 Model fit . . . 30
1.5.2 Preferences . . . 31
1.5.3 Utility differences by origins and destinations . . . 32
1.6 Counterfactual analysis . . . 34
1.6.1 Role of joint migration disutility . . . 35
1.6.2 Role of destination characteristics . . . 37
1.7 Conclusion . . . 41

Appendices **43**
1.A Figures . . . 43

2 Urbanization and Interethnic Marriages in Sub-Saharan Africa **47**
2.1 Introduction . . . 48
2.2 Conceptual framework: The role of urbanization for interethnic marriages 50
2.3 Data and descriptive evidence . . . 52
2.3.1 DHS . . . 52
2.3.2 GHS-SMOD . . . 52
2.3.3 Interethnic marriages across countries and over time . . . 54
2.3.4 Pooled sample statistics . . . 55
2.4 Estimating the rural-urban gap in intermarriage . . . 58
2.4.1 Empirical strategy . . . 58
2.4.2 Results . . . 58

2.4.3 Robustness checks . . . 60
2.5 Exploring channels: A matching model approach . . . 61
2.5.1 Set up . . . 63
2.5.2 Data structure and variables . . . 64
2.5.3 Estimation . . . 66
2.5.4 Counterfactual analysis . . . 70
2.6 Conclusion . . . 74

Appendices **76**
2.A Tables . . . 76
2.B Figures . . . 78
2.C Algorithms . . . 80

3 Attitudes towards Immigration in a Highly Multicultural Society: The Roles of Foreign Background and Local Exposure **81**
3.1 Introduction . . . 82
3.2 Data sources and stylized facts . . . 85
3.3 Empirical strategy . . . 92
3.4 Results . . . 94
3.5 Conclusion . . . 105

Appendices **106**
3.A Sample selection and representativeness . . . 106
3.B Summary statistics . . . 108
3.C Additional results . . . 108
3.C.1 Effect of migration background: Results for controls at the zip-code level . . . 108
3.C.2 Effects of individual level characteristics on attitudes towards immigration . . . 111
3.C.3 Heterogeneity Analysis: Interaction with individual level characteristics . . . 111
3.C.4 Bootstrapped standard errors . . . 120
3.D Pro-immigration index . . . 121
3.E Exposure to immigration at the workplace . . . 121

Bibliography **125**

List of Figures

1.1 Migration frequency by age and gender . . . 17
1.2 Share of couples from the same birthplace . . . 22
1.3 Distance between spouses' birthplaces . . . 22
1.4 Distance between spouses' birthplaces and current location . . . 24
1.5 Share of migrant couples by urban and rural residence . . . 25
1.6 Origin and destination labor force participation rate by type of move . 26
1.7 Actual versus predicted marriage rates . . . 30
1.8 Actual versus predicted matches by type of couple . . . 31
1.9 Effect of reduced joint migrant disutility on mixed marriages by location 36
1.10 Effect of reduced joint migrant disutility on marriage choices by origin 37
1.11 Singles at origin by policy scenario (rural origins) . . . 38
1.12 Singles at origin by policy scenario (urban origins) . . . 39
1.13 Effect of destination characteristics on joint and mixed marriages by location . . . 40
1.14 Effect of urban destination on marriage choices by origin . . . 41
1.15 Effect of high labor force participation on marriage choices by origin . 42
1.A.1 Urbanization trends in Indonesia and South-East Asia . . . 43
1.A.2 Selected provinces in the Indonesia Family Life Survey . . . 44
1.A.3 Distribution of age at marriage for men and women . . . 44
1.A.4 Effect of partially reduced joint migrant disutility on marriage choices by origin . . . 45
1.A.4 Effect of partially reduced joint migrant disutility on marriage choices by origin . . . 46

2.3.1 Country-level correlations between urban population and intermarriage share . . . 54
2.3.2 Average percent of interethnic marriages in Sub-Saharan Africa. . . . 55
2.3.3 Urbanization and Intermarriage over marriage cohorts . . . 56
2.3.4 Intermarriage by urban classification over marriage cohorts . . . 56
2.5.1 Marriage shares by education and ethnic group . . . 67
2.5.2 Shares of marriages by spouses' traits . . . 68
2.5.3 Education . . . 68
2.5.4 Ethnicity . . . 68
2.5.5 Intermarriage shares under baseline and counterfactual scenarios . . 72
2.B.1 Model Fit: Observed vs. predicted matching . . . 78

3.2.1 Attitudes toward immigration: Responses to the national survey . . . 86
3.2.2 Pro-immigration attitudes: Differences across groups . . . 87
3.2.3 Computation of the neighborhood composition . . . 89
3.2.4 Distribution of the neighborhood composition measures . . . 91

3.4.1 Immigration and willingness to avoid a neighbor of a certain ethno-racial type . . . 100
3.4.2 Robustness to the definition of the close neighborhood Effect of immigration on NTM . . . 102
3.A.1 Differences in covariate distribution in weighted, unweighted and consent sample . . . 106
3.A.2 Selection into consent by covariates (unconditional mean) . . . 107
3.A.3 Selection into consent by attitude (conditional mean) . . . 108
3.C.1 OLS effects of individual level characteristics on attitudes . . . 111
3.C.2 Heterogeneity analysis: Interaction with age . . . 112
3.C.3 Heterogeneity Analysis: Interaction with education . . . 113
3.C.4 Heterogeneity analysis: Interaction with inactivity in the labor market 114
3.C.5 Heterogeneity analysis: Interaction with origin (Portuguese) . . . 115
3.C.6 Heterogeneity analysis: Interaction with origin (FR-BE-DE) . . . 116
3.C.7 Heterogeneity analysis: Interaction with origin (Other EU) . . . 117
3.C.8 Heterogeneity analysis: Interaction with origin (Non-EU) . . . 118
3.C.9 Heterogeneity analysis: Interaction with origin (Second generation immigrant) . . . 119
3.D.1 Pro-immigration index . . . 122

List of Tables

1.1 Labor force participation rates across regencies . . . 19
1.2 Summary statistics . . . 21
1.3 Actual versus predicted population . . . 30
1.4 Estimation results . . . 32
1.5 Migration disutilities . . . 33

2.3.1 Sample summary . . . 53
2.3.2 Summary statistics (Pooled DHS couple data) . . . 57
2.4.1 Determinants of interethnic marriage (Linear Probability Model) . . . 59
2.4.2 Robustness checks . . . 62
2.5.1 Summary of matching data . . . 66
2.5.2 Matching model estimation results . . . 69
2.5.3 Counterfactual matrix of men's types . . . 71
2.5.4 Counterfactual matrix of women's types . . . 71
2.5.5 Changes in urban marriage selection by type and gender (with rural marriage surplus) . . . 74
2.A.1 Summary of aggregated groups for matching . . . 76
2.A.2 Education contingency table . . . 77
2.A.3 Ethnicity contingency table . . . 77
2.A.4 Robustness of matching estimation to exclusion of migrants . . . 77

3.4.1 Effect of socio-demographic characteristics and migration background (OLS) . . . 95
3.4.2 Effects of neighborhood exposure to immigration . . . 98
3.4.3 Heterogeneity along the cultural vs. economic dimensions – Effect of neighborhood exposure on NTM (OLS estimates) . . . 104
3.B.1 Summary statistics . . . 109
3.C.1 Effect of zip-code level characteristics on pro-immigration attitudes (OLS) . . . 110
3.C.2 Effect on NTM with alternative numbers of bootstrap replications . . . 120
3.E.1 Effects of firm exposure to immigration . . . 123

Abstract

This dissertation consists of three self-contained chapters centered around the economics of migration and marriage. Migration and marriage decisions are arguably two of the most consequential choices individuals make over their lifetimes. This thesis explores how these choices interact, and how structural and demographic changes related to migration can affect preferences and attitudes.

The first two chapters focus on the role of location in the formation of marriages. These chapters build on the framework introduced by Becker (1973). His model of the marriage market essentially rests on the premise that men and women form unions based on the utility they derive from the match, and that an equilibrium emerges from competition for spouses with the desired traits. This formalization of the economics of marriage spurred a new field of study, at large concerned with the determinants of marriage utility, the role of assortative mating, and the reconciliation of the theoretical predictions with empirical observations (e.g., Anderberg, 2004; Hamilton & Siow, 2007; Lam, 1988; Wong, 2003). With the seminal contribution of Choo and Siow (2006), who introduced random utility to identify the underlying preference structure from observed matches, estimation became possible within a two-sided discrete-choice framework. In the first two chapters, I use recent advancements of this work based on the estimation of general linear models introduced in Galichon and Salanié (2024).

Chapter 1 examines how location preferences enter into the marriage decision, with a focus on differences between rural and urban locations and between economic activity. The context of this study is Indonesia, a country with large internal migration flows and a high prevalence of marriage. Using data on migration histories and couples' joint location, I first show descriptive evidence that both men and women frequently move away from their birthplace to their spouse's home, and that migration distances can be large. Joint migration is equally observed, both to rural and urban destinations. I then estimate utilities of matching and staying at home, migrating together, and marrying someone from another market. Internal migration and marriage choices are connected within a Choo and Siow (2006) type matching model that allows for preferences over spousal and location types as introduced in Dupuy (2021). I find that couples migrating together face utility losses compared to those staying at home, and that urban destinations are typically preferred by migrants, both joint and independent. By simulating a marriage market in which joint migration becomes less costly, I quantify the substitution between joint migration and marrying at the destination. The results show that mixed marriages in urban locations are most responsive, primarily driven by rural couples forming at home and migrating together as opposed to independent migration. On the other hand, restricting joint migration increases mixed marriages in urban destinations by up to 17 percentage points.

Chapter 2 explores how urban and rural location interacts with ethnicity preferences

in Sub-Saharan Africa. We test the hypothesis that urban environments encourage interethnic marriages, through the composition of ethnic groups and increased intergroup contact. First, using cross-country data from Demographic Health Surveys and linking them to urbanization at the local level using standardized geodata, we compare marriages within the same locality at different levels of urbanization over time. We find a robust positive effect of urbanization on intermarriage, which increases by the degree of urbanization. In particular, the probability that a couple is interethnic is 13.6 percentage points higher in the highest density urban center compared to the least dense rural area. Next, we explore potential channels of this rural-urban gap, quantifying the role of ethnic preferences and group distributions. We consider a (closed) rural or urban marriage market where men and women trade off spousal education and ethnicity. The estimation of preference parameters in our data shows that, while matching on ethnic identity increases marital utility more than education matching for both rural and urban couples, this difference is smaller in urban couples. A potential explanation could be that, relatively, education becomes more important as a trait in marriage in urban settings. Further, our counterfactual simulations reveal that differences in the intermarriage rate between rural and urban markets can be in large part explained by ethnic preferences, compared to differences in group distributions. This can be seen as evidence that urban environments change preferences, e.g., through increased openness and the role of social norms, rather than merely affecting the availability of spouses from a certain group. This chapter's findings imply that intermarriage, commonly considered an indicator of social cohesion, can be influenced by structural changes such as urbanization.

Chapter 3 addresses demographic changes in form of immigration in high-income countries and how these relate to social cohesion measured by attitudes towards immigration. In particular, we study the factors that shape attitudes in Luxembourg, with a focus on local exposure to immigrants and individuals' own migration background. The setting of Luxembourg invites this type of analysis, given the country's long history of immigration and exceptionally large proportion of the population with foreign background. We combine an original survey collecting data on attitudes in 2021 with rich administrative data that helps us identify respondents' exposure to immigrants in their neighborhood. Overall, a large majority of respondents acknowledge that immigration enriches national identity and has a positive impact on the economy. Pro-immigration attitudes are stronger among second-generation immigrants and even more pronounced among first-generation immigrants from both EU and non-EU countries. In addition, we do not find that the total share of immigrants in the immediate neighborhood significantly influences attitudes towards the role of immigrants for the economy and the identity of the country. However, local exposure to recent immigration and to immigrants from non-EU countries correlates negatively with natives' perceptions of the current level of immigration. These findings suggest that embracing diversity and multiculturalism requires time for mutual adjustment by newcomers and their host community.

Declaration of Co-Authorship

Chapter 1: Marriage Matters: Internal Migration and Marital Sorting in Indonesia

This chapter is single-authored.

Chapter 2: Urbanization and Interethnic Marriage in Sub-Saharan Africa

This chapter is based on joint work with Camila Pineda Leon (University of Luxembourg) and Michel Tenikue (LISER).

My contributions to this chapter within the research team include conceptualization, methodology, data curation, software (implementation), formal analysis, visualization, validation, and writing (original draft, review & editing), according to the Contributor Role Taxonomy (NISO CRediT Working Group, 2022). All technical aspects of the paper were discussed and jointly decided by the three co-authors.

Chapter 3: Attitudes towards Immigration in a Highly Multicultural Society: The Roles of Foreign Background and Local Exposure

This chapter is based on joint work with Frederic Docquier (LISER), Michel Tenikue (LISER), and Aleksa Uljarevic (University of Luxembourg).

My contributions to this chapter within the research team include data curation, software (implementation), formal analysis, visualization, validation, and writing (review & editing), according to the Contributor Role Taxonomy (NISO CRediT Working Group, 2022). All technical aspects of the paper were discussed and jointly decided by the four co-authors.

Declaration of Use of Generative AI

The text, images, and tables in this thesis are my own (unless otherwise specified). Generative AI tools have only been used in accordance with the guidelines of the University of Luxembourg. I take full responsibility for the content of the thesis.

At early stages of my literature review, ResearchRabbit was used to search relevant literature as a complement to searching traditional databases. Microsoft Copilot and ChatGPT have been used for debugging Stata, Matlab, Python and Latex code. They were further employed for clarifying some theoretical concepts and methodology used in this thesis, together with verification from other sources such as peer-reviewed papers. Finally, these tools supported the editing and proofreading of individual sentences or small portions of text, primarily to suggest alternative phrasing or clearer expression. All outputs generated through these tools were critically reviewed, verified, and thoroughly revised by the author.

Chapter 1

Marriage Matters: Internal Migration and Marital Sorting in Indonesia

1.1 Introduction

With approximately 740 million internal migrants globally, mobility within a country's borders is a common phenomenon in many developing countries (UNDP, 2009). In Indonesia, Jakarta and other metropolitan centers are primary destinations for young migrants in search for better work and educational opportunities (Wajdi et al., 2017). Yet, sizable migration flows are also recorded across rural areas, highlighting economic pull factors in both rural and urban sectors (Bryan & Morten, 2019). For policy makers, these large mobility flows present a dual challenge of incentivizing migration to reap productivity gains, while avoiding overpopulation of developed regions through agglomeration forces (Gollin et al., 2014, 2021; Lagakos et al., 2023). A key aspect in designing appropriate policy responses is to understand what drives destination choices and the motives of migration.

In this chapter, I provide evidence that the decision (where) to migrate is closely interlinked with the decision (where) to marry. As Figure 1.1 shows, the majority of internal moves by men and women occur around the ages 20-30, coinciding with the timing of marriage for most people. Thus, location and marriage decisions may be interlinked in several ways. For example, Imbert et al. (2023) show that migrants may trade off destinations based on wages and amenities depending on whether they can bring their family or not. When it is costly to move together, migrants leave their family behind and choose higher wages over better living conditions. On the other hand, better labor market conditions elsewhere may incentivize marriage with a migrant when individual migration barriers exist (Amirapu et al., 2022). Lastly, finding a spouse in a different location may be a way to access better economic opportunities.[1] The aim of this chapter is to bring these potential choices together in a framework where men and women make their marriage decision with their location preferences in mind. I explore how men and women choose their spouse based on their location, and how joint migrants choose their destination. Further, I test how the decision to marry someone at the destination is affected by incentives or barriers to joint migrants.

I start by documenting several descriptive facts about couples' origins and locations. Using data from the Indonesia Family Life Survey (IFLS), I show that spouses primarily come from different subdistricts and distances between birthplaces can be large. There are important differences between urban and rural couples, with urban couples being generally more likely to have migrant spouses. Further, important interactions arise in the choice of rural and urban destinations in terms of economic activity, proxied by the average labor force participation rate: while rural-to-rural migrants (both joint and independent marriage migrants) on average live in locations with a higher labor force participation rate than their origin, rural-to-urban migrants are more likely to choose a destination that has a lower average labor force participation rate as their origin. This indicates important trade-offs between urban labor markets and those with high economic activity.

How do location preferences enter the marriage decision, and how much do couples value their location? To investigate this, I introduce a structural model of the marriage market where men and women can find a spouse in their home or across markets, while matching along location-specific traits and education. Thereby, couples can either be native (i.e., matching in the same location and stay), joint migrants, or mixed couples,

[1] For example, Rosenzweig and Stark (1989) show that marriage migration can serve as a mechanism for consumption smoothing for rural families in India.

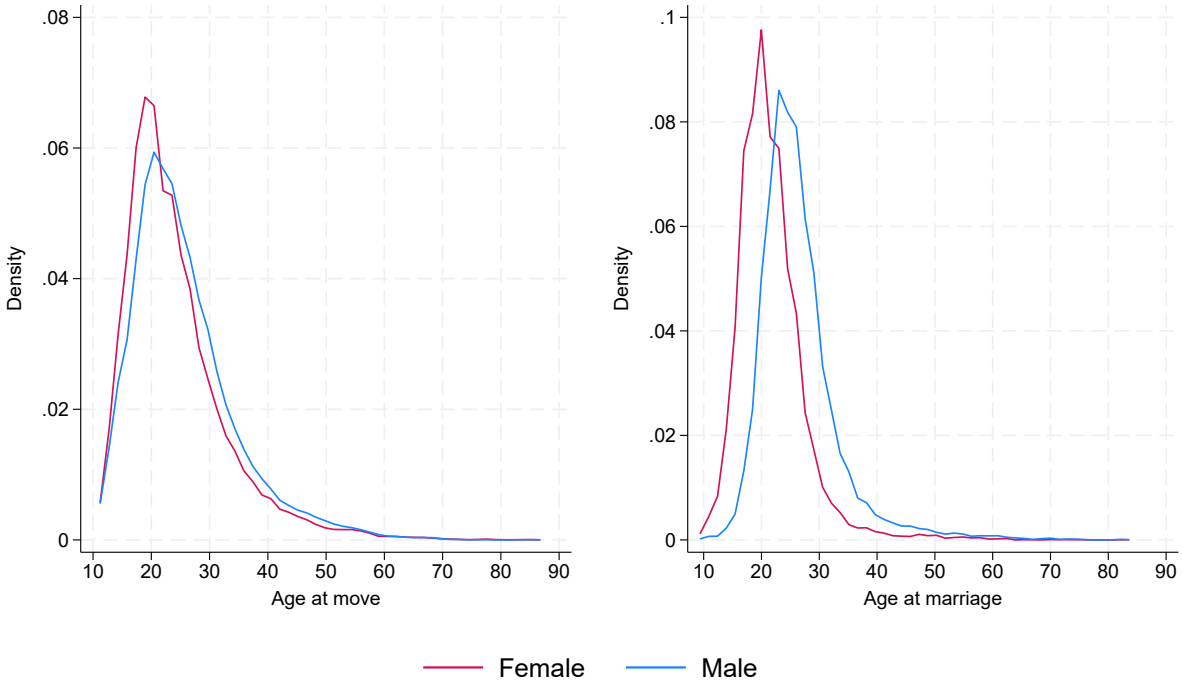


Figure 1.1: Migration frequency by age and gender

Notes: Data from Indonesia Family Life Survey (2014).

where one of the spouses is a migrant and one is a native. In particular, individuals or couples can move between rural and urban locations which are further distinguished by their labor force participation rate. From the observed matches, I estimate preferences based on a parametric joint utility function within a matching model with transferable utility.

The results show that the highest utility comes from native couples in any location. Couples migrating together to urban destinations receive positive, but lower utility levels and rural migration is associated with negative utilities. In other words, couples migrating together face utility losses compared to those staying at home. Similar destination-dependent results are recorded for mixed couples, where urban mixed couples receive higher utility than those in rural areas. This is true both for couples with a male or a female migrant. Further, matching on education plays an important role for all types of couples.

These findings are consistent with strong preferences for matching at the origin. The utility losses incurred by joint migrants may incorporate migration costs and the weakening of social ties that can benefit couples staying at home (Debray et al., 2025). The results further indicate that urban locations are preferred by both joint migrants and mixed couples. Given these findings, I turn to assessing how the "origin matching" preference interacts with migration preferences to urban destinations or destinations with a bigger labor market. To do this, I simulate counterfactual policy scenarios where (1) joint migration becomes less costly in utility terms, or (2) joint migration to some destinations is restricted or too costly. The resulting equilibrium matching into mixed couples indicates how responsive this migration channel is to changes in the availability of joint migration.

Reducing the utility loss from joint migration to all destinations affects mixed

marriages primarily in urban areas, with both male and female migrants from all origins. This serves as evidence that matching with someone at the origin would be preferred if joint migration was less costly. In light of migration frictions for couples, marriage migration can serve as a way to realize migration preferences, particularly to urban areas. This is further confirmed when restricting urban access for joint migrants, which increases mixed marriages there by 7-17 percentage points, depending on the labor force participation rate. By contrast, restricting joint access to destinations with higher labor force participation, regardless if rural or urban, only affects mixed marriages in urban areas. This suggests that marriage migration only compensates for having a spouse from the same origin when the destination is urban.

Related Literature This chapter relates to several strands of the literature. First, it contributes to the understanding of assortative mating based on the origin, and in particular how location background enters marriage utility. The assortative mating literature has primarily focused on socioeconomic traits, including education and income (Anderberg et al., 2019; Charles et al., 2013; Eika et al., 2019; Pesando, 2021), or socially ascribed characteristics such as ethnicity, race, and religion (Bandyopadhyay & Green, 2021; Crespin-Boucaud, 2020; Goldman et al., 2025). Such studies typically find strong positive matching along these dimensions. Where people are born and grow up can be highly correlated with such background characteristics and may itself constitute a basis of matching. For example, couples that come from the same region may share similar worldviews or aspirations. My findings present evidence that, indeed, men and women sort positively into marriages based on the same origin.

I further show that migration preferences also enter the marriage decision. With this, my study aligns with other works focusing on marriage migration. Several studies have investigated marriage across borders in Asia and Europe (Adda et al., 2025; Ahn, 2021; Farahzadi, 2024; Kawaguchi & Lee, 2017; Weiss et al., 2018). These accounts highlight the role of sex imbalances, income differences, and residence incentives. In a similar vein, I test how location-based economic incentives are taken into account in the marriage decision when there are no borders, but potentially other (social or economic) frictions. Internal marriage migration has been brought forward as an economic strategy, though mostly as a coping mechanism after income shocks (Becerra-Valbuena & Millock, 2021; Gray & Mueller, 2012; Hidrobo et al., 2022; Rosenzweig & Stark, 1989). Closest to my approach are Amirapu et al. (2022) and Dupuy (2021), which incorporate the role of urban labor markets into the marriage decision. With respect to their findings, I elaborate on the substitutability between joint migration and marriage migration.

Lastly, I add to the research on frictions in rural-to-urban migration by taking into account the role of marriage markets. With persistent gaps in productivity and living standards between rural and urban areas in many developing countries (Gollin et al., 2014; Lagakos, 2020), previous studies have sought to explain these discrepancies by migration costs (Bryan et al., 2014; Lagakos et al., 2020, 2023). On the other hand, Imbert and Papp (2020) show that potential migrants trade off higher wages in the city with work opportunities in their village, indicating that they disproportionately value staying at home or have a strong distaste for urban living conditions. Indeed, Bryan and Morten (2019) find that both migration costs and amenity differences are important drivers of productivity gaps. To what extent utility from migration may differ between independent and joint migrants has not been part of this research agenda, with the exception of Imbert et al. (2023). They show that the decision to bring family

can lead to different trade-offs between wages and amenities for migrants compared to those leaving family behind. Compared to their approach, I consider the option of migrating together or migrating alone and marrying someone at the destination. Thereby, I explore a different type of trade-off between home and destination marriage markets and how they relate to the (dis)utility of joint migration.

Context Indonesia has been rapidly urbanizing in the last decades, reflecting large-scale rural-to-urban movements across the country.[2] Several urban centers have developed primarily across the main islands (UNDESA, 2019). Urban areas are characterized by a mix of industries, including trade, services and manufacturing, while rural areas are mainly agricultural. As census data from 2010 shows in Table 1.1, these sectoral differences also imply differences in labor force participation and wage employment. While urban regencies (the second-highest administrative level) are on average more than 1.5 larger in terms of their working-age population (15 and older), they exhibit an average labor force participation rate that is almost 10 percentage points lower than that observed in the rural population. On the contrary, the mean share of wage employment in the labor force population is twice as high in urban compared to rural areas. This implies that a high share of the rural labor force works self-employed or as a unpaid (family) worker. Further, both rural and urban areas experience large variation in labor force participation and wage employment.

Table 1.1: Labor force participation rates across regencies

	Total		Rural		Urban	
	Mean	SD	Mean	SD	Mean	SD
Population size	114,493.941	93193.716	86,835.547	73403.296	142,804.684	102349.058
in labor force	0.696	0.460	0.743	0.437	0.650	0.477
in wage employment	0.403	0.490	0.275	0.446	0.547	0.498

Notes: Data from 493 regencies in IPUMS census 2010. The table displays the average working-age population size (aged 15 and older) in a regency, share of the working-age population in the labor force, and share of the labor force in wage employment.

These geographical differences in labor markets may be potential push or pull factors for movements within the country. This is in line with large migration flows within rural or urban locations in Indonesia as found in Bryan and Morten (2019). In this study, I therefore focus on moves both across rural and urban categories as well as across locations with differing labor force participation rates. Compared to wages as a measure of productivity, this indicator has the advantage of taking into account economic activity beyond wage employment. Further, while wage employment of women has been increasing over time, they are to a large extent involved in non-wage employment. Focusing on wages only may therefore understate the relevance of female labor force participation compared to men (Schaner & Das, 2016).

The remainder of this chapter is structured as follows. In section 1.2 I describe the data sources and the sample, in particular with respect to couples' origins. Section 1.3 introduces the matching model. Section 1.4 describes the estimation approach, with the results summarized in 1.5. In section 1.6, I introduce the counterfactual analysis. Lastly, 1.7 concludes.

[2] As Figure 1.A.1 in Appendix 1.A shows, Indonesia surpassed the regional average for urbanization in the early 1990s, reaching an urban population of over 50% by 2010.

1.2 Data

The primary source of data is the 2014 round of the Indonesia Family Life Survey (IFLS). The IFLS is a multi-purpose panel survey that was administered to households and individuals over five waves (1993-2014). The first wave of the survey was launched in 1993, where households in 13 of the 27 Indonesian provinces were interviewed, see Figure 1.A.2.[3] The sampling scheme stratified on provinces and urban/rural location, resulting in a sample representative of 83% of the population (Strauss et al., 2016). In the subsequent rounds, target households and respondents were re-interviewed and, if necessary, tracked to their new residence where new household members were also interviewed. The residents roster links each member to their spouse in the household. For respondents above the age of 15, detailed migration histories are collected. For all panel and new respondents in 2014, I use information on their birthplace, residence at age 12, and their current residence. While this information is available on the sub-district (kecamatan) level, i.e., the third-level adminstrative subdivision, I can also identify if their community at each point in time is rural or urban. I further use information on urban and rural labor force participation by regency, i.e., the second-level administrative unit, from the IPUMS sample of the 2010 census (Ruggles et al., 2024).

1.2.1 Sample

My population of interest is, primarily, married (or co-habitating) individuals living in the same household. Further, for the estimation I will use information on single, never-married men and women.[4] I focus my analysis on men aged 29 to 67, and women aged 24 to 63. As shown in Figure 1.A.3, the median age at marriage is 25 for men and 21 for women, with 75% of men and women married by the age of 29 and 24, respectively. Therefore, I use this as a threshold for the age at which most men and women are married in expectation. This allows me to infer the utility for independent migration, without the expectation of getting married at the destination, from single migration. The descriptive findings are not substantially altered by this age restriction. Married individuals are matched with their spouse living in the same household, and singles are observed in their current residence.[5]

Table 1.2 describes important characteristics of the sample. The average age of the sample is 43 for men and 38 for women, with around 95% of men and women being married or living with their spouse. Almost 60% of men and women live in urban locations, while only 28% of men and 33% of women were born in an urban town or city. This reflects the urbanization trends of the country, and is underlined further by the fact that around 68% of men and 66% of women have moved away from their birthplace in their lifetime. In terms of education, men and women have similar levels of schooling. The majority of men and women in the sample have a secondary degree,

[3] The remaining provinces were excluded due to remoteness and political risk.

[4] I exclude individuals that are divorced, separated, or widowed. Of all respondents aged 15 or older, 4.6% are widowed and 2% are divorced or separated. Shares of these groups are similar for men and women, and when taking into account older cohorts.

[5] There is only one case where one man is matched to two spouses in the same household. I drop this household altogether. The marriage history questionnaire reveals that around 4% of men indicate having more than one current wife. In this case, I treat the spouse living in the same household as the "main wife", which is included in the sample.

Table 1.2: Summary statistics

	Men		Women	
	Mean	SD	Mean	SD
Age in years	42.978	10.051	38.169	10.233
Married or cohabitating	0.948	0.221	0.947	0.223
Live in urban	0.571	0.495	0.593	0.491
Born in urban	0.284	0.451	0.325	0.469
Ever moved	0.675	0.468	0.664	0.472
Eduaction				
No education	0.024	0.153	0.035	0.184
Primary education	0.322	0.467	0.294	0.456
Secondary education	0.505	0.500	0.501	0.500
Tertiary education	0.150	0.357	0.170	0.376
Observations	8466		8462	

Notes: The sample contains men aged 29-67 and women aged 25-63. The variable "born in urban" is defined as the place of birth being a town or city. The variable "ever moved" is a binary variable that is 1 if the respondent does not live in the same sub-district as they were born or if they have indicated a move of over 6 months.

with around 15% of men and 17% of women achieving higher than secondary education. Only around 2% of men and women have no education, and around a third of men and women have achieved a primary degree.

1.2.2 Descriptive evidence

This section explores the composition of couples in terms of migrants and natives, the distances between spouses' birthplaces, and salient differences between rural and urban couples. Further, it discusses the types of destinations chosen by joint migrant couples and male or female independent (marriage) migrants.

Figure 1.2 shows the shares of couples from the same province, regency and sub-district in urban and rural Indonesia. Provinces are the highest administrative level in the country, covering large geographic areas and a number of regencies. Generally, a greater share of rural couples was born in the same location than is the case for urban couples. While the majority of couples is from the same province, the share drops to around 51% of urban couples and 68% of rural couples from the same regency. Only around half of rural and one third of urban couples were born in the same sub-district. Marriages across sub-districts can reach large distances, as seen in Figure 1.3. Consistent with the implications of Figure 1.2, around 50% of marriages are within a 5km distance.[6] Still, there is large variation in distances of cross-district marriages, with some reaching 800km and more.

[6] The distances in Figure 1.3 are calculated using geo-reference of the birth sub-districts. This implies that I cannot distinguish distances of marriages within sub-districts, which may be sizable in some areas given the high variation of sub-district sizes across the country.

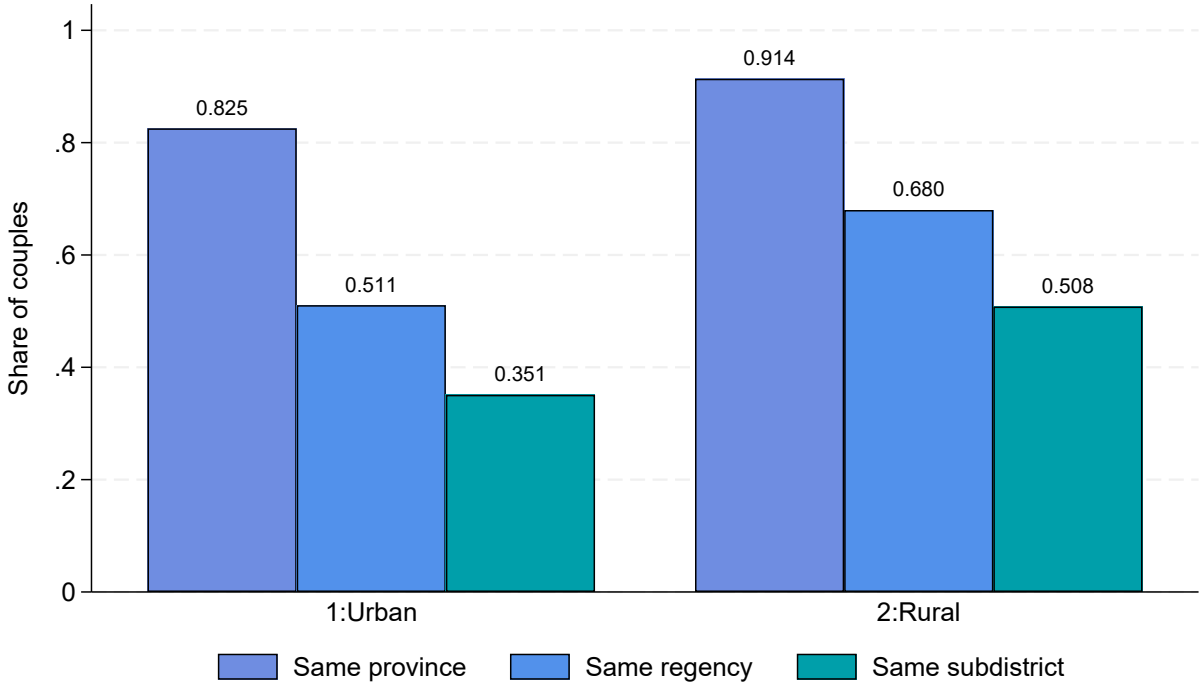


Figure 1.2: Share of couples from the same birthplace

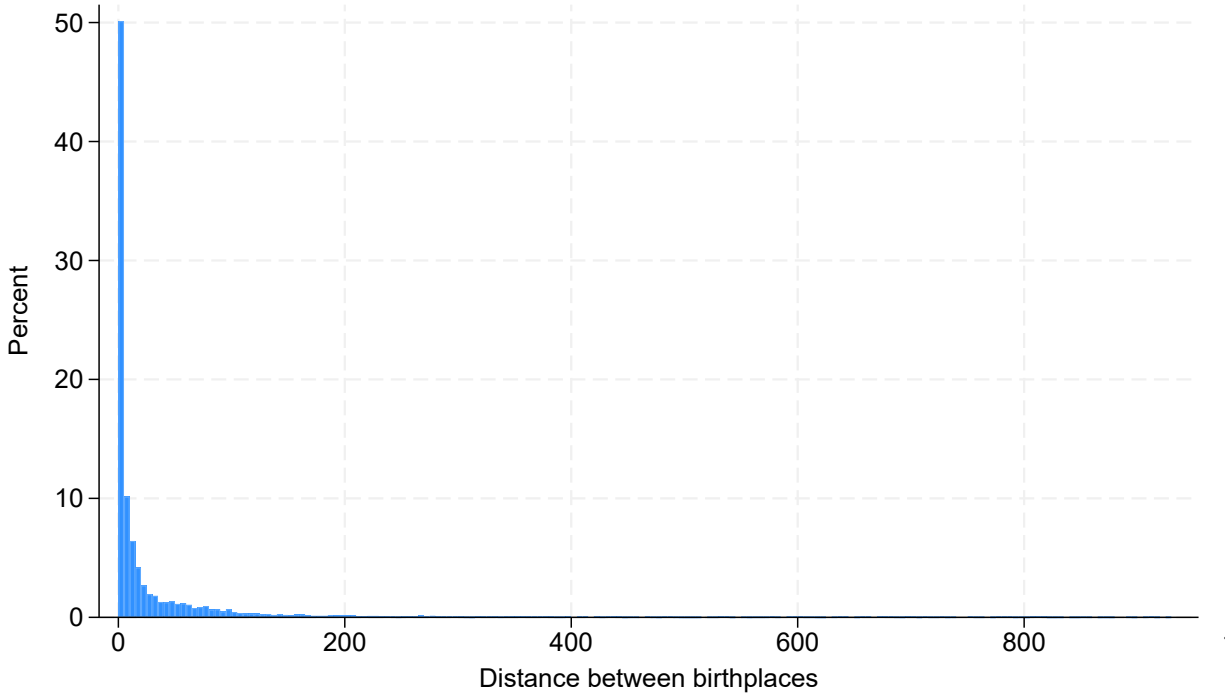


Figure 1.3: Distance between spouses' birthplaces

Notes: Distances are calculated using longitude and latitude of the sub-district at birth. Bins represent 5km distances. Observations with distances larger than the 99th percentile are dropped as outliers.

Next, I investigate in more detail the composition of couples in terms of their

birthplaces and where they live now. Figure 1.4 indicates three main types of couples: (1) those with both spouses coming from close to their current residence (observations in the bottom left corner of the plot), (2) couples that move together from the same birthplace (observations on the diagonal), and (3) couples where one spouse is from close to their current residence and the other has moved there (observations along the x- and y-axis). Again, the data suggests that while there is a large share of couples living close to their birthplace, this is far from the only option. Notably, couples with both male and female (independent) migrants are prevalent. Further, all types of migrants, regardless if joint, male, or female, seem to cover extensive distances. This may reflect different locality norms present in Indonesia. Specifically, some ethnic groups primarily practice matrilocality, while others observe patrilocal (or neolocal) norms (Bau, 2021).[7] Locality traditions govern the post-marital location of the couple, which may be close to the husband's (in case of patrilocality) or the wife's (matrilocality) kin, or in an entirely new household (neolocal). Thereby, these traditions are intrinsically linked to migration behavior of couples.

Figure 1.5 shows the differences of types of couples by urban and rural residence. Couples that jointly move across sub-districts make up around 9% of urban couples and 10% of rural couples. Around 42% of urban couples consist of a migrant wife and 40% have a migrant husband. By contrast, the share of rural couples with a female migrant is 31% and those with a male migrant is 27%. Thus, while similar shares of couples moving together exist in rural and urban areas, urban couples are largely characterized by independent migrants.

Lastly, Figure 1.6 shows the type of rural and urban destinations different migrant types choose in terms of labor force participation.[8] It indicates that rural-to-rural moves are primarily taking place to destinations with higher labor force participation than the origin, while rural-to-urban moves are more likely to go in the opposite direction. This pattern appears for couples that move jointly as well as those where either the husband or the wife has moved. However, Figure 1.5 showed that the share of mixed marriages is higher in urban areas compared to rural areas. In combination, this could imply that marriage migrants are more prone to enter urban destinations by making concessions on the size of the labor market, while for joint migrants the trade-off is less clear.

While these descriptive facts shed light on the observed choices in partners and locations, they cannot give insight into the underlying preference structure that leads to these observations. For example, from observing mixed couples in urban locations, we cannot distinguish if these couples formed because of intrinsic preferences, or as an equilibrium outcome from bounds to joint migration. To shed light on how these choices interact with each other, we move to a structural model in the next section.

[7] Ethnicity and religion play a major role in the family formation. This is reflected in the 2010 census, where 89% of couples are from the same ethnic group.

[8] The figure displays moves across regencies instead of sub-districts as more reliable labor force participation rates are taken from the 2010 census, which is only available on the regency level. The trends are replicable when taking information on labor force participation from the IFLS and plotting moves across sub-districts.

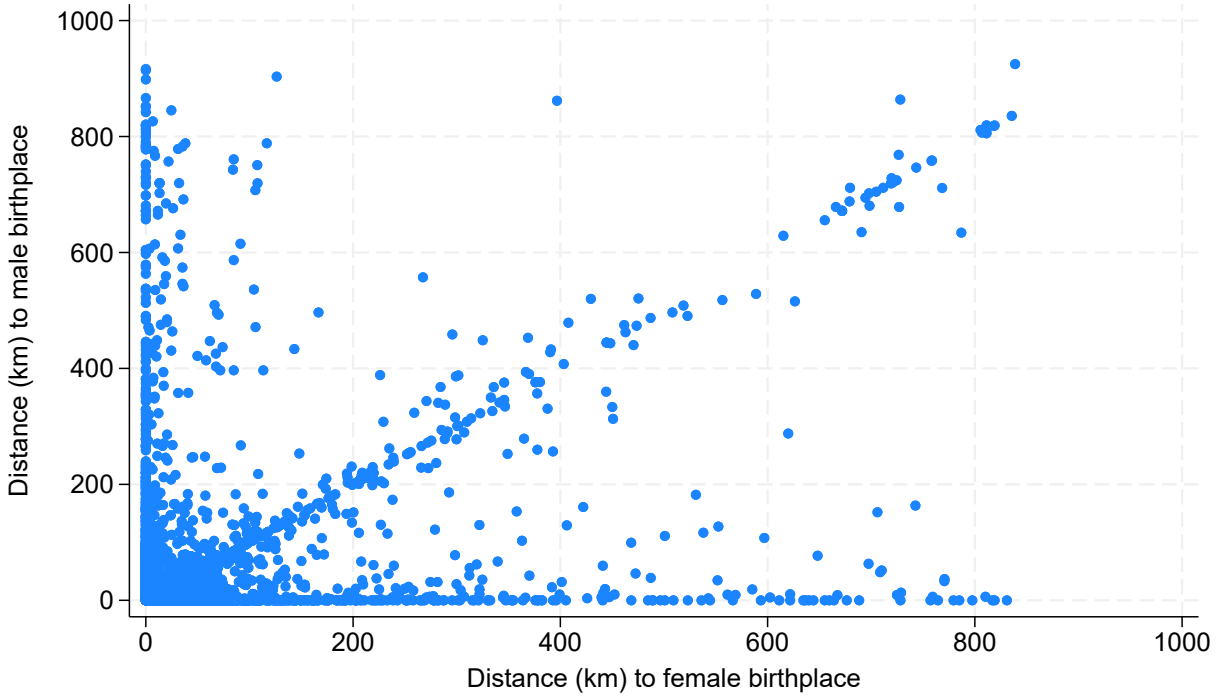


Figure 1.4: Distance between spouses' birthplaces and current location

Notes: Distances are calculated using longitude and latitude of the sub-district at birth and current residence. Observations with distances larger than the 99th percentile are dropped as outliers.

1.3 Model

This section introduces a two-sided matching model with transferrable utility including location choice to explain how spousal characteristics and location characteristics enter the marriage decision. This model allows for a general equilibrium analysis taking into account preferences of both sides of the marriage market and the supply of potential spouses. The estimation of structural parameters from the model further enables the application of a counterfactual analysis in section 1.6.

The decision-makers in this market are men and women, who match given their personal preferences and bargaining based on utility transfers. It should be noted that while arranged marriage traditionally has played a role in the family formation in Indonesia, the decision-making has shifted from parents to the spouses in the recent decades, with almost 95% of couples recorded in the IFLS and married after 2000 confirming they were the ones who chose their spouse.[9] Still, even in the case of family involvement in the match, I will assume that the decision was made based on the spouses' utility.

Transfers can be in form of monetary or non-monetary exchanges. For example, some ethnic groups in Indonesia engage in bride price customs (Ashraf et al., 2020). However, these customs often include the exchange of gifts that are hard to be priced accurately. Beyond that, utility transfers may include other non-monetary concessions such as child rearing or housework. Thus, while transfers are an equilibrium object in

[9] This is compared to 61% of couples that were married in 1970 or earlier.

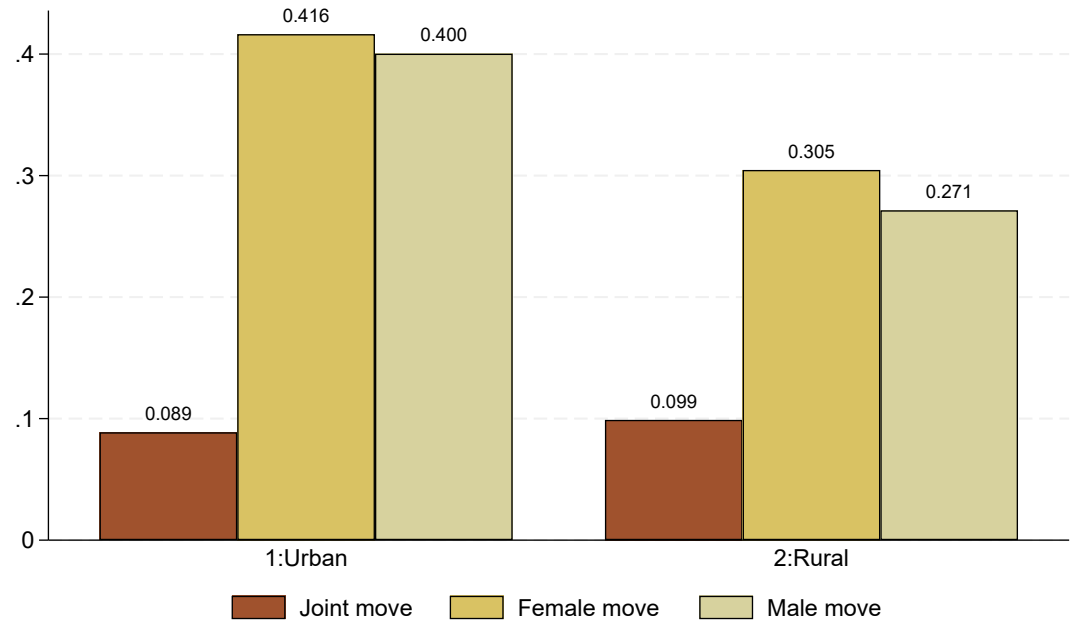


Figure 1.5: Share of migrant couples by urban and rural residence

Notes: Migration is defined as moves across sub-districts from birthplace to current residence. The graph shows the shares of couples in rural and urban locations of each migrant type: joint move (both spouses from same sub-district), female move (wife from different sub-district), male move (husband from different sub-district).

the model, they will not be explicitly identified.

The model assumes a frictionless market as in Choo and Siow (2006). This implies that individuals have perfect information about the distribution of the types of potential partners in different marriage markets. Given that in my model, people care about their partners' origin (in terms of rural/urban status and level of labor force participation) and their education level, it may be plausible to assume perfect information and the absence of search costs. Still, marrying someone away from home may involve overcoming physical and social constraints, e.g., leaving family and social ties behind. While these potential frictions cannot be directly measured, I intend to shed light on "utility costs" involved in joint or independent migration through the counterfactual exercise.

Lastly, I introduce location as a choice within the marriage decision following Dupuy (2021). As the model is static, the marriage and migration decision is made once and is irreversible. For marriage, this implies no divorce and re-marriage is possible. For the location choice, this can be interpreted as permanent migration.

1.3.1 Two-sided maximization problem

The market consists of men of type $x \in \mathcal{X}$ and women of type $y \in \mathcal{Y}$, which includes the origin, denoted by $Z(x)$ and $Z(y)$, as well as education. There are finite masses of types $N(x)$ and $M(y)$ in the market. Formally, men and women choose their partner from the choice sets $\mathcal{Y}_0 = \mathcal{Y} \cup \{0\}$ and $\mathcal{X}_0 = \mathcal{X} \cup \{0\}$, respectively, where 0 denotes the option to remain single. Locations are of type $z \in \mathcal{Z}$, including the origin $Z(x)$ for men and $Z(y)$ for women.

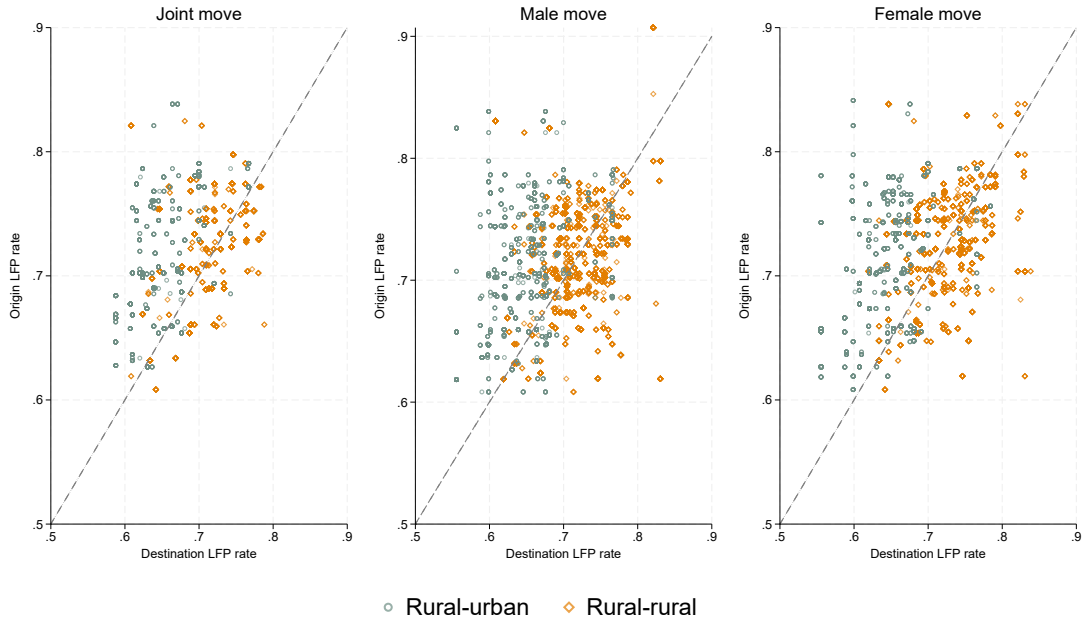


Figure 1.6: Origin and destination labor force participation rate by type of move

Notes: Graphs plot the origin and destination labor force participation rate on the regency level. Migration is defined by moves across regencies for each migrant type: joint move (both spouses from same sub-district), female move (wife from different sub-district), male move (husband from different sub-district).

The utility of man i of type x married to a woman of type y living in location z is given by:

$$u_i(x) = \alpha(x, y, z) + t(x, y, z) + \varepsilon_i(y, z)$$

Similarly, the utility of a y-type woman j married to a man of type x in location z is:

$$v_j(y) = \gamma(x, y, z) - t(x, y, z) + \eta_j(x, z)$$

where $\varepsilon_i(y, z)$ and $\eta_j(x, z)$ are idiosyncratic tastes of man i (woman j) for a type of woman y (man x), drawn from an Extreme Value type I distribution.

The systematic part of their utilities, $\{\alpha(x, y, z)+t(x, y, z)\}$ and $\{\gamma(x, y, z)-t(x, y, z)\}$, consists of an intrinsic utility, $\alpha(x, y, z)$ or $\gamma(x, y, z)$, and a utility transfer, $t(x, y, z)$, from one side of the match to the other. The intrinsic (or direct) utilities account for utility that is received from matching with a man of type x or a woman of type y as well as the utility of living in location z. The transfer allows for a bidding process and connects the utilities of i and j in case they match. If i and j match, they receive a joint systematic utility

$$\begin{aligned}\Phi(x, y, x) &= \alpha(x, y, z) + t(x, y, z) + \gamma(x, y, z) - t(x, y, z) \\ &= \alpha(x, y, z) + \gamma(x, y, z)\end{aligned}$$

where the transfer $t(x, y, z)$ cancels out.[10]

If man i and woman j decide to stay single, they receive a reserve utility of

$$u_i^0(x) = \alpha(x, 0, z) + \varepsilon_i(0, z)$$

and

$$v_j^0(y) = \gamma(0, y, z) + \eta_j(0, z)$$

Man i and woman j then maximize their utilities according to:

$$\max_{y \in \mathcal{Y}^0 z \in \mathcal{Z}} \{\Phi(x, y, z) - \gamma(x, y, z) + \varepsilon_i(y, z), \quad \alpha(x, 0, z) + \varepsilon_i(0, z)\}$$

for man i, and

$$\max_{x \in \mathcal{X}^0 z \in \mathcal{Z}} \{\Phi(x, y, z) - \alpha(x, y, z) + \eta_j(x, z), \quad \gamma(0, y, z) + \eta_j(0, z)\}$$

for woman j.

This two-sided maximization problem produces a matching vector μ in equilibrium, which contains the masses of couples of types x and y in location z, $\mu(x, y, z)$, as well as the masses of single men of type x, $\mu(x, 0, z)$ and the masses of single women of type y, $\mu(0, y, z)$. Matching μ must satisfy the feasibility constraints:

$$N(x) = \sum_{yz} \mu(x, y, z) + \sum_{z} \mu(x, 0, z), \quad \text{and} \quad M(y) = \sum_{xz} \mu(x, y, z) + \sum_{z} \mu(0, y, z)$$

Therefore, man i and woman j face several trade-offs. First, given their preferences for their partners' type and the masses of men or women of this type in market z, they decide on where to get married to whom. They further trade off the utility from a potential match with the outside option of staying single. Lastly, they take into account the utility they receive as a single native or migrant in location z, with the utility they receive as a native, migrant, or mixed couple in location z. Given the finite types of men and women in the market, both sides negotiate an adequate transfer $t(x, y, z)$.

1.3.2 Equilibrium

Given the distributional assumptions on the idiosyncratic tastes, the maximization problem above can be solved as a two-sided discrete choice problem. As outlined in Dupuy (2021), this results in the following solutions:

$$\Phi(x, y, z) = log\left(\frac{\mu^2(x, y, z)}{\mu(x, 0)\mu(0, y)}\right) \tag{1.1}$$

for couples, where $\Phi(x, y, z)$ is the joint systematic utility from $\alpha(x, y, z) + t(x, y, z) + \gamma(x, y, z) - t(x, y, z)$ and $\mu(x, 0)$ and $\mu(0, y)$ are the masses of singles of types x or y at origin location $Z(x)$ or $Z(y)$, respectively. As I do not observe data on the utility transfers $t(x, y, z)$, for couples I will identify the joint surplus $\Phi(x, y, z)$ from marriage in location z compared to staying single at home. For singles, the solutions are:

[10] This result is conveniently used in the identification of the joint utility $\Phi(x, y, z)$ instead of separate utilities $\alpha(x, y, z)$ and $\gamma(x, y, z)$, as the (hypothetical) transfer between husbands and wife is not observed in the data.

$$\alpha(x,0,z) = log\left(\frac{\mu(x,0,z)}{\mu(x,0)}\right) \tag{1.2}$$

$$\gamma(0,x,z) = log\left(\frac{\mu(0,z,x)}{\mu(0,y)}\right) \tag{1.3}$$

Assuming that we observe a stable equilibrium in the data, these result imply that the objects $\Phi(x,y,z), \alpha(x,0,z)$ and $\gamma(0,x,z)$ are a direct mapping of the observed choices through the matching pattern $\mu = \{\mu(x,y,z), \mu(x,0,z), \mu(0,y,z)\}$.[11]

1.3.3 Identification

I will leverage these results to identify several key objects of interest:

(1) Native couples' joint utility in location $z = Z(x) = Z(y)$, i.e., the utility of a couple with the same origin and still residing in the origin;
(2) Joint migrant couples' utility in $z \neq Z(x) = Z(y)$, i.e., the utility of a couple with the same origin in a new location;
(3) Mixed couples' utility in location $z = Z(x) \neq Z(y)$ or $z = Z(y) \neq Z(x)$, i.e., the utility of a couple with one migrant spouse; and
(4) Single migrants' utility in $z \neq Z(x)$ or $z \neq Z(y)$.

By normalizing the utility of single natives to 0, i.e., $\alpha(x,0,Z(x)) = 0$ and $\gamma(0,y,Z(y)) = 0$, the above utilities are expressed as the surplus of a (native/migrant/mixed) couple or migrant single of type x and y compared to being single at home.

1.3.4 Spousal and location types

With the model, I aim to primarily estimate who marries whom in terms of their origin and where different types of couples settle. To do this, I distinguish the location types between rural and urban, as well as the level of labor force participation, which can be either low or high compared to a threshold value. Given that the potential origin types $Z(x)$ and $Z(y)$ are identical to the location types z, the set of origin and location types are summarized as $z = Z(x) = Z(y) = \{RL, RH, UL, UH\}$, where R is rural, U is urban, L is low and H is high labor force participation. In terms of data, the rural/urban status is taken from the IFLS.[12] The low or high labor force participation rate is determined using information from the 2010 census and linked on the regency-level to the origin and current location. The threshold is set at the country-level average (69.6% as indicated in Table 1.1). Therefore, migration is determined by changes in the type of location from the type of origin, i.e., across rural/urban classification and/or across levels of labor force participation.

Another attribute that may be important in determining both a marital match and the propensity to migrate is the education level. Therefore, men and women are further

[11] The equilibrium matching is stable under the conditions that (1) no married individual would rather be single (no divorce), and (2) no two individuals who are not married to each other would rather be married with each other. The first condition is in line with the low number of observed divorces in the sample and the high stigma surrounding divorce in the country. While it is hard to justify the second condition with data, it is a standard and intuitive assumption when analyzing marriage markets in general.

[12] While the rural/urban status is collected directly for the community in which the respondent is found, for the origin I take self-reported information on the birthplace being a village, small town, or big city. I consider villages rural and small towns and big cities as urban.

distinguished by their schooling, which can be either low – primary level or below (P), or high – high school level or above (HS). Assortative mating, especially for higher levels of schooling, has been documented for developed and developing countries (Eika et al., 2019; Esteve et al., 2016; Pesando, 2021). Data from the 2010 Indonesian census shows that around 72% of couples had the same education level. However, this may be due to the distribution of educational attainment of men and women or preferences. Further, differences in assortative mating may arise across different types of locations given higher potential incomes.

1.4 Estimation

To estimate preferences over types of spouses and locations, I assume that the joint surplus of couples and the migration surplus of singles are linear in parameters:

$$\Phi^{\lambda}_{xyz} = \phi'_{xyz}\lambda = (\phi_{xy} \times z)'\lambda \tag{1.4}$$

$$\alpha^{\delta}_{x0z} = \Lambda'_{x0z}\delta = (\Lambda_{x0} \times z)'\delta \tag{1.5}$$

$$\gamma^{\tau}_{0yz} = \Gamma'_{0yz}\tau = (\Gamma_{0y} \times z)'\tau \tag{1.6}$$

where ϕ'_{xyz} contains indicator variables determining a marital match between natives, joint migrants, and mixed couples as well as couples with the same level of education for each location z. Similarly, Λ'_{x0z} and Γ'_{0yz} contain indicator variables for single migrants to z, as well as an interaction term with higher education.

Using the system of equations (1.1), (1.2), and (1.3), I then employ a moment-based estimation strategy by Poisson regression following Galichon and Salanié (2024).[13] To do this, we can combine the expressions with the feasibility constraints and rearrange them into a Poisson model of the following form:

$$\mathbb{E}[\mu^0_{xyz} \mid \phi'_{xyz}, \Lambda'_{x0z}, \Gamma'_{0yz}, \mathbf{I}_X, \mathbf{I}_Y, w] = \exp\left(w\phi'_{xyz}\lambda + w\Lambda'_{x0z}\delta + w\Gamma'_{0yz}\tau + w\mathbf{I}_X + w\mathbf{I}_Y\right) \tag{1.7}$$

where $\mu^0_{xyz} = \{\mu_{xyz}, \mu_{x0z}, \mu_{0yz}\}$, $\mathbf{I}_X$ and $\mathbf{I}_Y$ are indicator variables for types x and y and w are weights that are 0.5 for couples and 1 for singles. In other words, I run a Poisson regression of μ^0_{xyz} on the matching indicators in ϕ'_{xyz}, Γ'_{0yz} and Λ'_{x0z} with weighted x- and y- fixed effects. These fixed effects control for the total masses of types of x and y in the market according to the feasibility constraints of the model. Practically, this means that we can control for the supply of men and women by their education level and origin characteristics. The singles' migration parameters δ and τ then quantify the importance of an attribute (here: the education level) for the selection into single migration (by location), and the couples' parameters collected in λ quantify the importance of matching on a particular attribute (origin and education) by location.

[13] This method is a fast and straightforward alternative to the moment-matching procedure introduced in Galichon and Salanié (2022). The goal of this procedure is to match moments in the data to simulated moments in the model, using the given set of basis functions.

1.5 Estimation results

1.5.1 Model fit

Before presenting the estimation results, this subsection provides evidence that the predictions from the applied specification fit the data well. Table 1.3 shows the actual and predicted population sizes in each location, in total, by gender, and by type of household. Overall, the predicted data matches well the observed data, both in terms of total population sizes and disaggregated populations of interest. This is true for all locations, with no major divergences.

Table 1.3: Actual versus predicted population

	Total		Men		Women		Couples		Migrants	
	Data	Pred.	Data	Pred.	Data	Pred.	Data	Pred.	Data	Pred.
RL	1971	2146.58	980	1078.53	991	1068.05	939	1002.28	358	456.61
RH	4182	4290.24	2108	2149.10	2074	2141.13	2020	2068.40	535	608.52
UL	4948	4676.43	2485	2351.26	2463	2325.18	2303	2195.81	1807	1692.94
UH	3264	3251.75	1646	1640.11	1618	1611.64	1534	1502.05	1559	1528.39

Notes: The table shows actual number of matches (count data) from the IFLS 2014 data and those predicted by the model. The column "Total" provides the total population. "RL" stands for rural, low economic activity, "RH" for rural, high economic activity, "UL" for urban, low economic activity, and "UH" for urban, high economic activity locations.

The fit also holds when considering the marriage shares by genders as in Figure 1.7. The marriage shares of both men and women are slightly underpredicted in rural areas, but differences are small. For urban marriages, the data is well represented.

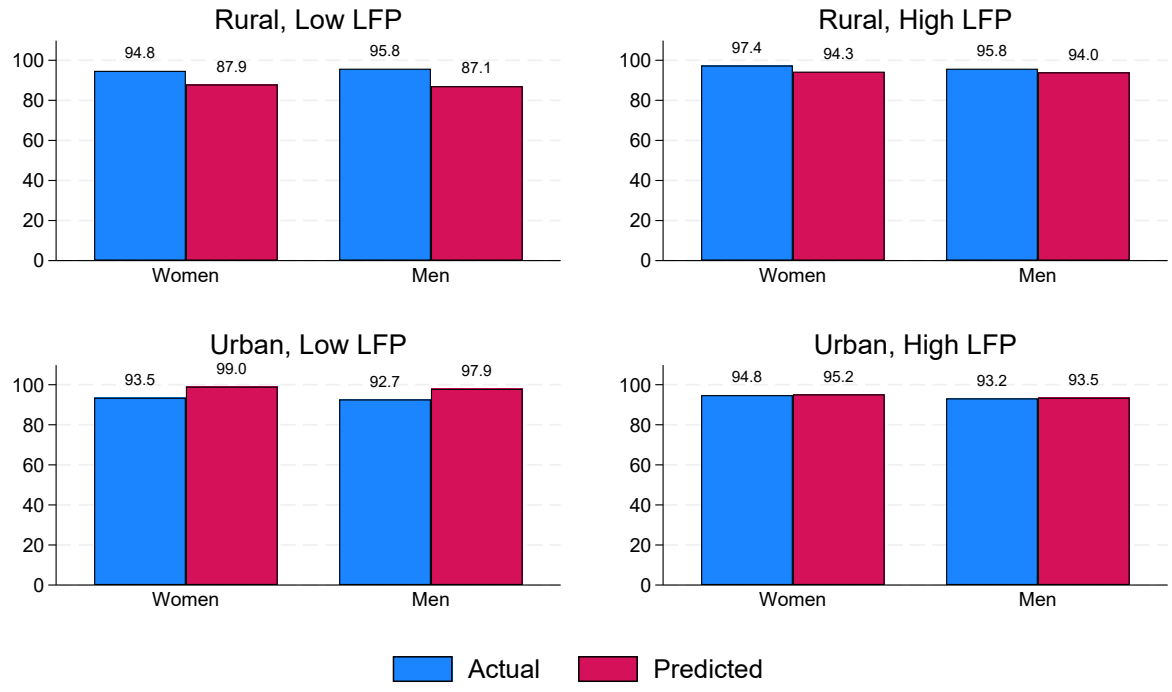


Figure 1.7: Actual versus predicted marriage rates

Lastly, we can assess the fit of the predictions in terms of type of couples observed. The main categories of interest are displayed in Figure 1.8. Again, predictions are close to the shares observed in the data. Joint migrant couples are slighly overpredicted in rural locations and underpredicted in urban locations. However, these differences are small.

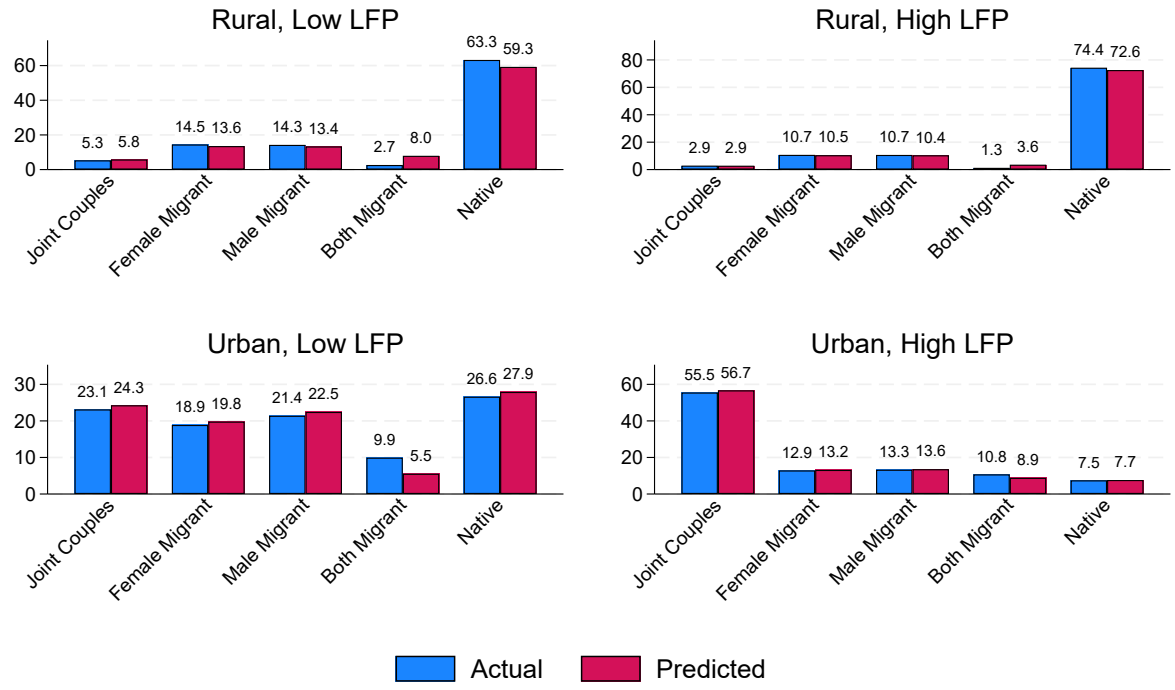


Figure 1.8: Actual versus predicted matches by type of couple

1.5.2 Preferences

The first set of results pertain to utilities received by a type of match or single. As described above, the parameters can be interpreted as a surplus from marriage or single migration with respect to staying single at home (given a specific type). They can also be seen as preference parameters in comparing different choices with each other. Table 1.4 presents the main parameter estimates for the model specified in equation (1.7) above. It shows that native couples receive positive marriage utility across locations, which is consistently higher than joint migrants and mixed couples. On average, joint migrants in rural destinations receive negative utility, while those in urban areas receive a positive level that is still lower than for natives. With similar levels of positive assortative mating in education for native and migrant couples, only those rural migrant couples matching on their education level increase their utility to a positive level. Taken together, these results imply a strong preference both for matching with someone from the same origin and for staying there together.

Turning to mixed marriages, the results are similar for couples with a male or female native. Mixed marriages in urban locations receive higher utility than those in rural areas. While for couples with a native husband there is no large distinction between locations with higher or lower economic activity, those with a native wife seem to receive higher utility in locations where labor force participation is lower. This could

Table 1.4: Estimation results

	Rural		Urban	
	Low LFP	High LFP	Low LFP	High LFP
Both native	4.023	4.72	5.817	2.985
	(0.377)	(0.344)	(0.719)	(0.399)
Same education	2.296	2.207	2.87	2.956
	(0.337)	(0.217)	(0.199)	(0.194)
Joint migrants	-2.93	-2.461	1.726	2.473
	(0.408)	(0.436)	(0.715)	(0.568)
Same education	2.592	3.451	2.058	2.245
	(0.592)	(0.671)	(0.954)	(0.749)
Wife native	-.149	.242	2.102	.388
	(0.659)	(0.503)	(0.897)	(0.622)
Same education	1.707	1.947	2.998	3.004
	(0.718)	(0.620)	(0.755)	(0.791)
Husband native	-.706	.538	2.143	1.475
	(0.550)	(0.583)	(0.816)	(0.705)
Same education	2.168	1.478	2.795	2.051
	(0.774)	(0.651)	(0.611)	(0.835)
Male single migrant	-2.932	-.605	-.603	-.532
	(0.988)	(1.541)	(1.236)	(1.070)
High education	.854	-.977	.544	.813
	(1.092)	(1.486)	(0.950)	(0.642)
Female single migrant	-1.99	-1.5	-.65	-.252
	(1.653)	(1.098)	(1.163)	(1.141)
High education	.426	-.149	1.197	.597
	(1.404)	(0.689)	(0.809)	(0.657)

Notes: The table presents the parameter estimates for different matching indicators for each location. Standard errors are presented in parenthesis below the estimates. LFP = labor force participation rate.

imply that men migrating to these types of locations benefit more from marrying a native than when migrating to urban destinations with better labor markets.

Lastly, migration utilities for singles can give insight into how men and women value the outside option of independent migration without marriage. While precise estimation is difficult due to the low number of observed single migrants, the results suggest that, similar to couples, urban destinations are preferred to rural ones. However, single migration is costly anywhere for both men and women, and this is only partly offset by a higher education level.

1.5.3 Utility differences by origins and destinations

The results in Table 1.4 indicate a utility loss from migrating both independently and as a couple, compared to staying at the origin. The results for joint migrants are an average of the utility received in each destination, without distinction between origins. To further investigate the utility differences between natives at the origin and joint migrants by destination, here I re-estimate equation (1.7) with a full set of origin-destination parameters for joint migrants. Thereby, I can calculate the utility

differences between a couple from each origin staying behind (native) and the same type of couple migrating to location z. Given that transfers are unobserved, this approach is only feasible for joint migrants. For mixed couples, I am unable to identify the spouses' individual utilities in the marriage, and thereby cannot pinpoint any disutility related to the migrant spouse's move. The results for single migrants can potentially give insight into the utility differences faced when moving independently, under the assumption that those that marry at the destination and those that stay single moved under the same conditions. However, as previously stated, singles' utility differences cannot be estimated precisely given the sample size. Instead, in the next section I will investigate how responsive mixed marriages are to changes in the migration disutility of joint migrants, to give insights on the relative utility costs between joint and marriage migration.

For joint migrant couples, the disutilities are calculated by taking the difference between the native parameter and the migration parameter by origin-destination pair:

$$\Delta_o^d = \lambda_o^N - \lambda_o^d$$

where Δ_o^d denotes the joint migration disutility, or utility loss, of a couple from origin o in destination d, the parameter λ_o^N indicates the utility from native couples by origin, and λ_o^d indicates the utility from joint migrants by origin-destination pairs. Due to data limitations, I restrict the possibility of migration from urban areas with high labor force participation to rural destinations with low labor force participation[14].

Table 1.5: Migration disutilities

	Location			
Couples' Origin	Rural, Low LFP	Rural, High LFP	Urban, Low LFP	Urban, High LFP
Rural, Low LFP	0.00	6.93	0.48	9.20
Rural, High LFP	7.34	0.00	5.54	1.28
Urban, Low LFP	5.01	10.13	0.00	6.13
Urban, High LFP	excl.	3.55	3.96	0.00

Notes: Disutilities are calculated as the difference between utility from migration to a destination and utility from staying at the origin.

Table 1.5 provides the results of this exercise. Several things can be noted for joint migrants. Generally, all moves are characterized by utility losses, with moves to rural destinations associated with higher disutilities on average. This reflects the results found in Table 1.4. Rural couples from origins with lower labor force participation face higher disutility from moving to urban destinations with higher compared to those with lower labor force participation. Similar results with respect to synergies in the labor market level apply for those from rural areas with higher labor force participation. While the latter might be expected, the former is more counter-intuitive. It implies that couples receive some utility from synergies in the level of the labor markets between the origin and destination, rather than simply from a higher level. These results seem to hold also for couples originating from urban markets. In the following counterfactual analysis, I therefore focus on overall joint migration costs as well as the roles of urban destinations versus those (rural or urban) with higher labor force participation.

[14] As the cells for joint migrants from this origin-destination pair are zero, the Poisson estimator excludes this parameter to ensure identification.

1.6 Counterfactual analysis

Recent evidence in the literature on internal migration has indicated that, rather than migration intentions, high migration costs can be a main deterrent of mobility (Bryan & Morten, 2019; Lagakos et al., 2020). Experimental studies have shown that even small monetary incentives can lead to an increase in migration (Bryan et al., 2014; Lagakos et al., 2023). However, these settings are typically focused on seasonal or temporary migration, where the household remains in the (often rural) origin and one (often male) householder leaves to the city for temporary work. It is more difficult to design an experiment that targets permanent relocation of families. Yet, these types of policies could have large effects on overall productivity (Bazzi et al., 2016). In the following, I explore the consequences on the marriage market of a policy that incentivizes mobility of couples disproportionately. These could be, for example, policies that target married couples specifically (such as the Indonesia transmigration program carried out between 1950 and 2000), or that offer support for dependent spouses or families of migrants.

On the other hand, governments may want to navigate migration flows to highly productive destinations and thereby control population dynamics. Typically, such policies aim at restricting access to public services for migrants, with the hukou system in China being a well-known example. Yet, many other countries follow similar practices (Bloom & Khanna, 2007). Such policies may affect joint migrants disproportionally, as they value amenities at the destination more than independent migrants (Imbert et al., 2023). Therefore, marriage migration may be a channel to circumvent these adverse conditions. A second counterfactual exercise will therefore test how deterrents to joint migration into urban or high-productivity locations affect marriage decisions at home and at the destination.

Following Galichon and Salanié (2022), the counterfactual equilibrium matching is computed using the results (1.1)-(1.3) from section 1.3.2 in an iterative projection fitting procedure (IPFP) algorithm. Rearranging the expressions and plugging them into the feasibility constraints (1.3.1) results in the following two expressions:

$$\mu_{x0} = \left(\sqrt{\frac{N_x}{\sum_{z\in\mathcal{Z}} L_{xz}} + \left(\frac{\sum_{y\in\mathcal{Y},z\in\mathcal{Z}} K_{xyz}\sqrt{\mu_{0y}}}{2\sum_{z\in\mathcal{Z}} L_{xz}} \right)^2} - \frac{\sum_{y\in\mathcal{Y},z\in\mathcal{Z}} K_{xyz}\sqrt{\mu_{0y}}}{2\sum_{z\in\mathcal{Z}} L_{xz}} \right)^2$$

$$\mu_{0y} = \left(\sqrt{\frac{M_y}{\sum_{z\in\mathcal{Z}} P_{yz}} + \left(\frac{\sum_{x\in\mathcal{X},z\in\mathcal{Z}} K_{xyz}\sqrt{\mu_{x0}}}{2\sum_{z\in\mathcal{Z}} P_{yz}} \right)^2} - \frac{\sum_{x\in\mathcal{Y},z\in\mathcal{Z}} K_{xyz}\sqrt{\mu_{x0}}}{2\sum_{z\in\mathcal{Z}} P_{yz}} \right)^2$$

with $K_{xyz} = exp(\frac{\Phi^{\lambda}_{xyz}}{2})$, $L_{xz} = exp(\alpha^{\delta}_{x0z})$, and $P_{yz} = exp(\gamma^{\tau}_{0yz})$. The parameters $\lambda, \delta,$ and τ are adjusted according to the counterfactual scenario tested, as explained below. A fixed-point algorithm (IPFP) is employed to find equilibrium expressions of (μ_{x0}, μ_{0y}). The steps of the algorithm are the following:

1. taking an initial guess of μ_{x0},
2. update values of μ_{0y} using current values of μ_{x0},
3. update the values of μ_{x0} using the current values of μ_{0y},
4. go back to step 2 until convergence.

Once a solution for the expressions (μ_{x0}, μ_{0y}) is achieved, we can find the new equilibrium matching patterns from:

$$\mu_{xyz} = K_{xyz}\sqrt{\mu_{x0}\mu_{0y}},$$

$$\mu_{x0z} = L_{xz}\mu_{x0},$$

and

$$\mu_{0yz} = P_{yz}\mu_{0y}$$

Comparing the matching patterns under observed and counterfactual utilities allows for the quantification of the substitutability of key choices of interest: staying and marrying at home, migrating together, or marrying elsewhere.

1.6.1 Role of joint migration disutility

To understand the role of joint migration on marriage market decisions, I simulate several policy scenarios in which the utility to jointly migrate becomes less "costly" in utility terms. For this, I use the concept of migration disutility established above and gradually add part of the native utility to that of migrants.

Recall that the baseline disutility from joint migration is $\Delta_o^d = \lambda_o^N - \lambda_o^d$. Then the counterfactual disutility $\Delta_o'^d$ is the baseline disutility reduced by a factor $x \in [0.25, 0.5, 0.75, 1]$ of the native utility:

$$\begin{aligned} \Delta_o'^d =& \Delta_o^d - x\lambda_o^N \\ \Delta_o'^d =& \lambda_o^N - (\lambda_o^d + x\lambda_o^N) \end{aligned} \tag{1.8}$$

Gradually adding part of the native utility (25, 50, 75 and 100%) allows me to explore the trajectory of trade-offs made given different levels of migration disutility. The extreme case where the entire native utility is added mimics a scenario where joint migrants receive the same utility as if they were staying at their origin, plus the utility they receive from a specific destination. However, from a policy perspective, it seems plausible that any interventions to incentivize internal migration would only partly offset the disutility faced by couples. For example, money transfers for migrants would ease the cost of migration, but would not affect the role of networks at the origin. On the other hand, facilitating networks for migrants at the destination would not fully offset other costs faced by joint migrants.

Figure 1.9 presents the results for mixed marriages from this exercise. It displays changes in mixed marriages with either a male or female migrant spouse in different locations, based on the extent of native utility added to joint migrants' utility. The share of mixed marriages in the scenario with 0% added native utility is equivalent to the baseline prediction. Mixed marriages at baseline make up between 10 and 20% of marriages depending on the location, with the highest share in urban areas with lower labor force participation. Generally, urban areas experience a steeper decline of mixed marriages when joint migration becomes more attractive, with a total drop to around 5%. This implies a decrease in the share of mixed couples by about half in the extreme case. By comparison, the decrease of mixed marriages is less steep in rural areas and only takes effect once joint migrants' utilities reach the maximum. The magnitudes are very similar for mixed couples with either a male or female migrant spouse. These results suggest that marriage migration to urban destinations is more

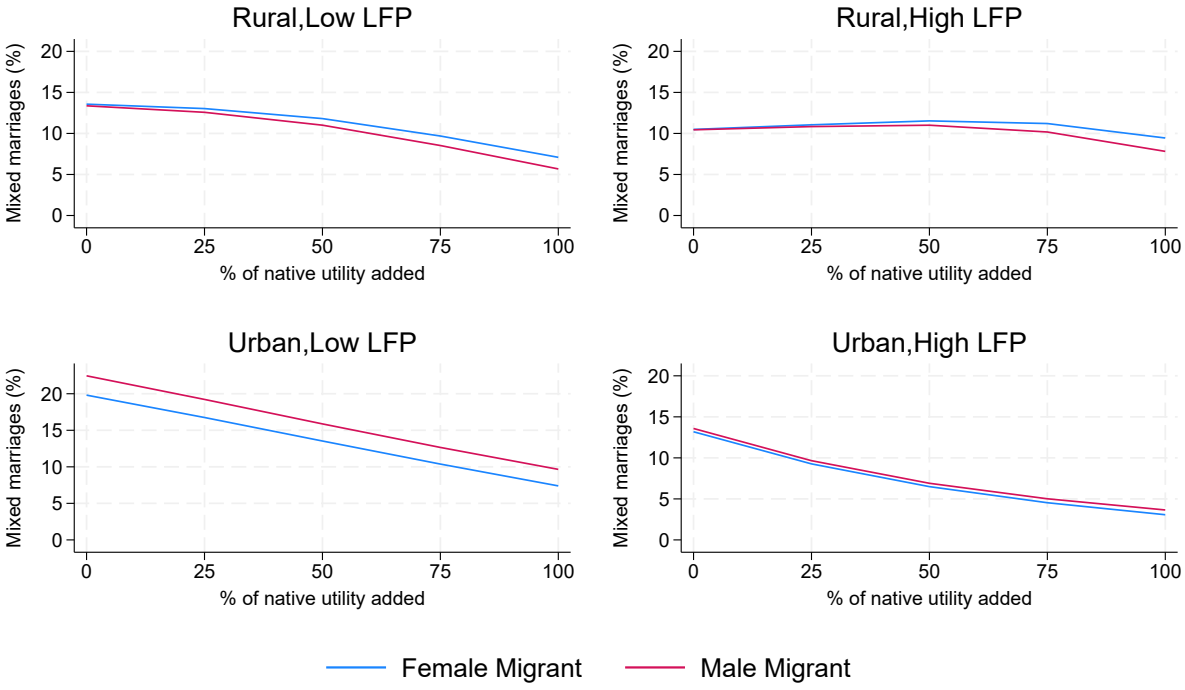


Figure 1.9: Effect of reduced joint migrant disutility on mixed marriages by location

Notes: Graphs depict mixed marriages with either a female or a male migrant, as a share of all marriages in a location. The shares are calculated for counterfactual scenarios where 25, 50, 75, or 100% of the native utility are added to joint migrants' utility, as outlined in equation 1.8.

responsive to changes in the utility for joint migration, especially for small benefits to joint migration. In other words, it seems that high disutility from joint migration leads to a large share of men and women entering the marriage market at the destination, especially in urban markets.

To further investigate which origins are primarily affected, Figure 1.10 depicts the effect of increasing the joint migrant utility (by 100% of native utility) on marriage choices in each origin.[15] The figure shows that the policy incentivizes joint migration from all origins, with a larger share of joint migrants forming in rural origins. These changes emerge both through reductions in the share of men and women migrating to marry away from home, and from couples forming at home that would have stayed at the origin. As the utility difference from staying at home and migrating together is now equalized, they are more likely to move away. Therefore, the downward trends of mixed marriages at the destination are not only due to a decrease in marriage migrants, but also due to new couples marrying at home and becoming joint migrants.

Lastly, we can observe how increasing utility from joint migration affects the outside option of staying single. As Figure 1.11 shows, for rural men and women this leads to a reduction in singlehood, both native and migrants. This means that even more couples form than before, as they now have the option to migrate together. The same results are seen in urban origins with lower labor force participation rates (Figure 1.12), albeit to a lesser extent and only at higher rates of utility added. Single migrants

[15] Figure 1.A.4 in Appendix 1.A presents the results for each partial utility added. It shows that the effects grow proportionally with the amount of utility added.

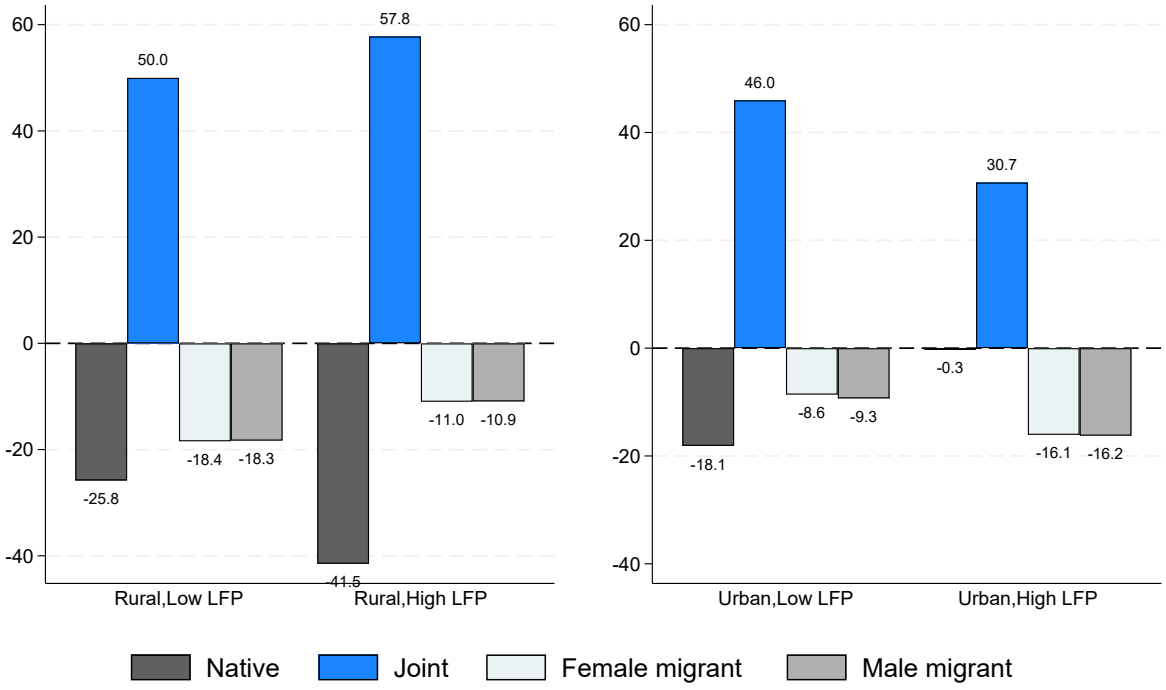


Figure 1.10: Effect of reduced joint migrant disutility on marriage choices by origin

Notes: The graph depicts the difference of the share of each category in the counterfactual to their baseline shares. The calculations are based on the 100% counterfactual scenario, where the entire native utility is added to joint migrants' utility.

from these locations are rare to begin with, and those that would stay only select into marriage once the joint migration utility is sufficiently high. Similarly, singles in urban locations with better labor markets are largely unaffected, likely because the gain from migrating with a spouse does not compensate the utility they receive from staying in their original labor market.

1.6.2 Role of destination characteristics

As shown in Figure 1.9, the type of destination plays a role in the responsiveness of mixed marriages to joint migration incentives. To investigate further how different traits – urban location or high economic activity – interact with the selection into joint- or marriage migration, I employ a second set of counterfactual exercises. In particular, I restrict joint migration as a channel to enter either urban destinations or those with high labor force participation, using the parameters for joint migration from origin o to destinations $d = \{UL, UH\}$ or $d = \{RH, UH\}$:

$$\lambda_o'^{UL} = \lambda_o'^{UH} = -\infty \tag{1.9}$$

$$\lambda_o'^{RH} = \lambda_o'^{UH} = -\infty \tag{1.10}$$

Thereby, we can quantify the "joint migration channel" to urban or high-economy destinations as compared to entering independently. Given that the previous findings

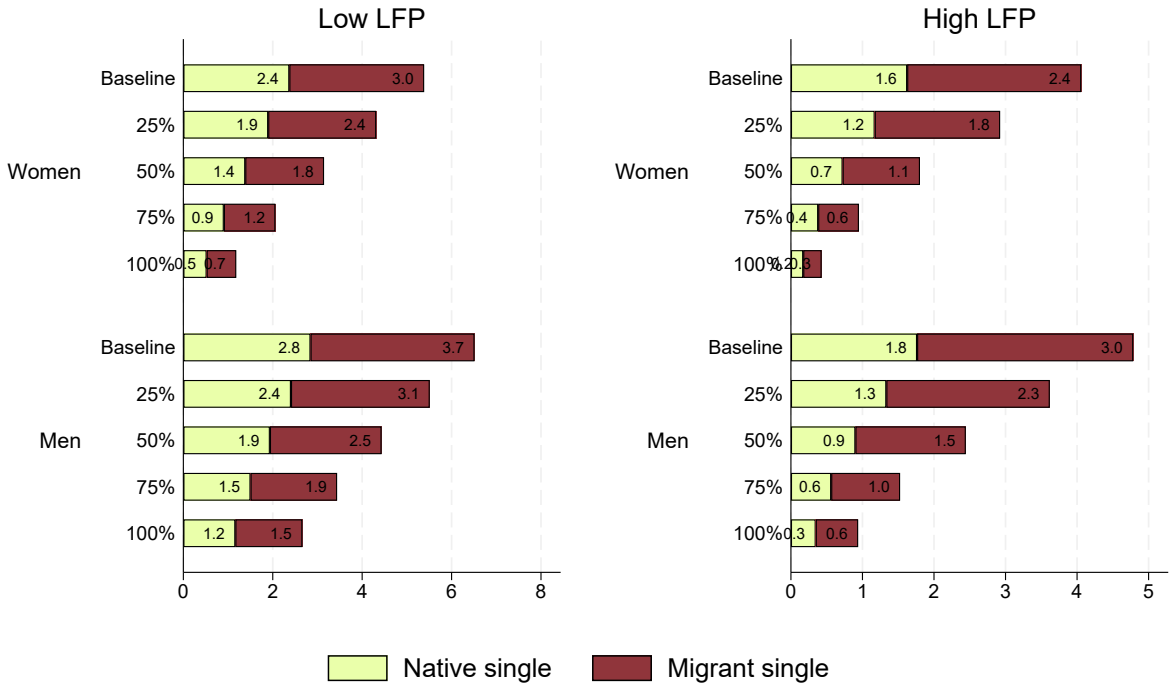


Figure 1.11: Singles at origin by policy scenario (rural origins)

Notes: The graph depicts the shares of single men and women that stay or migrate under baseline conditions and when joint migration disutility is decreased by 25, 50, 75 or 100% of the native marriage utility. LFP = labor force participation rate.

suggest a high preference for matching at the origin, once migration is restricted couples may be more inclined to form and stay at home. Alternatively, those who move together at baseline may choose a different destination or find a spouse at a preferred location. This boils down to a trade-off between matching at the origin and migrating to urban or high-economy destinations. Shutting down the option to migrate together to certain destination may trigger overall equilibrium effects in all locations.

Results of this counterfactual exercise are presented in Figure 1.13. The graph on the left relates to equation 1.9 and the one on the right to equation 1.10. They present the changes in the shares of joint migrant or mixed couples with a female or male migrant in the 4 locations: rural (R) or urban (U) with low (L) or high (H) labor force participation. By construction, joint migration to urban (in the left panel) and high-economy destinations (in the right panel) drops fully. When joint migration to urban destinations is restricted, this leads to a jump in mixed marriages by around 8 to 17 percentage points, depending on the level of the labor market. These effects are very similar for couples with a female or a male migrant. By contrast, joint migration to rural destinations is unaffected. In other words, there is no change in terms of destinations by migrant couples if they cannot enter urban areas. Rather, they do not form in the first place and men and women are more likely to enter the marriage market in urban destinations.

When restricting access to both rural and urban destinations with better labor markets (right panel), the drop in joint migrants is again compensated mainly by both male and female marriage migrants. However, this is only true for urban destinations,

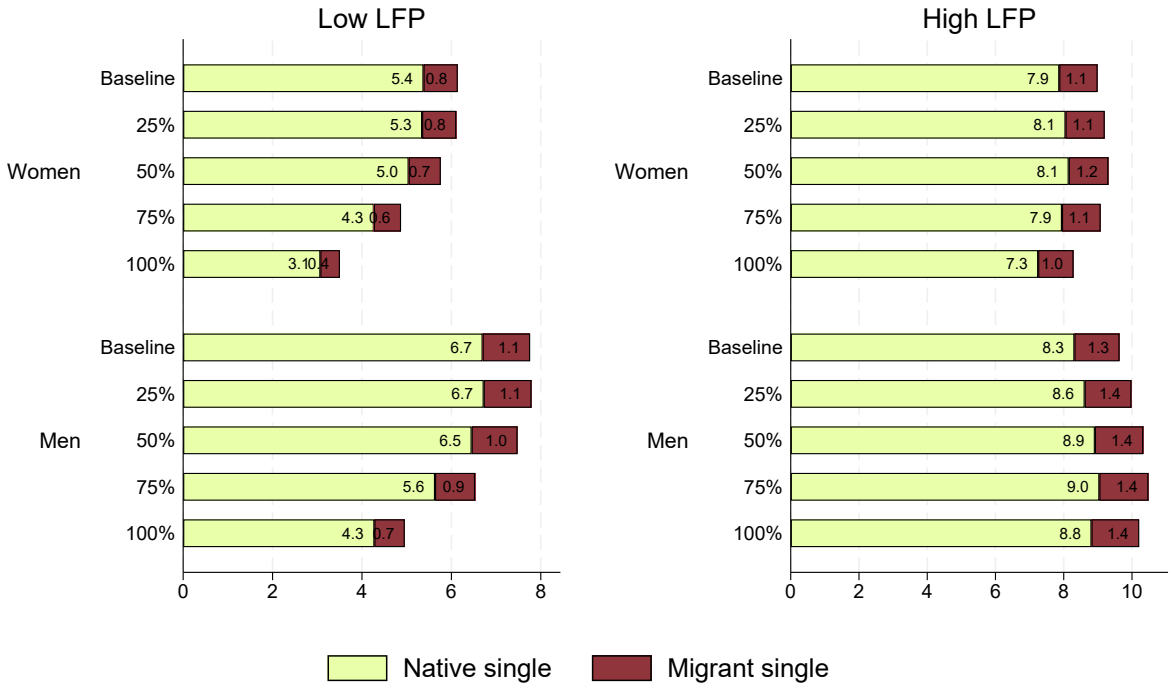


Figure 1.12: Singles at origin by policy scenario (urban origins)

Notes: The graph depicts the shares of single men and women that stay or migrate under baseline conditions and when joint migration disutility is decreased by 25, 50, 75 or 100% of the native marriage utility. LFP = labor force participation rate.

where the share of mixed marriage with a male or female migrant again rises by around 17 percentage points. On the contrary, restricting entry into rural high-type destinations does not lead to an increase in the share of mixed couples. If anything, the share of mixed couples slightly reduces, likely because the share of native couples not able to move to urban destinations now grows.

These results suggest that urban destinations, and in particular those with better labor markets, draw in more marriage migrants when joint migration is impossible. On the contrary, rural locations with higher labor force participation seem to be more attractive for joint migrants, and are not chosen through marriage migration as a substitute.

Figures 1.14 and 1.15 examine these choices from the perspective of origin. Starting with the scenario where joint migration to any urban destination is restricted, Figure 1.14 highlights how rural- and urban-born men and women change their marriage and migration choice compared to before. The majority of the changes observed in the destinations seem to stem from rural-born men and women. In particular, the share of couples matching and staying at home increases by around 15-22 percentage points, depending on the labor force participation rate at origin. Those from rural origins with better labor markets are more likely to stay. On the contrary, rural-born men and women from places with lower labor force participation are more likely to select into marriage migration as a response. In comparison, urban-born do not adjust their marriage choices by a lot, and only marginally increase their marriage migration. This implies that the increase in mixed marriages observed in urban areas is primarily

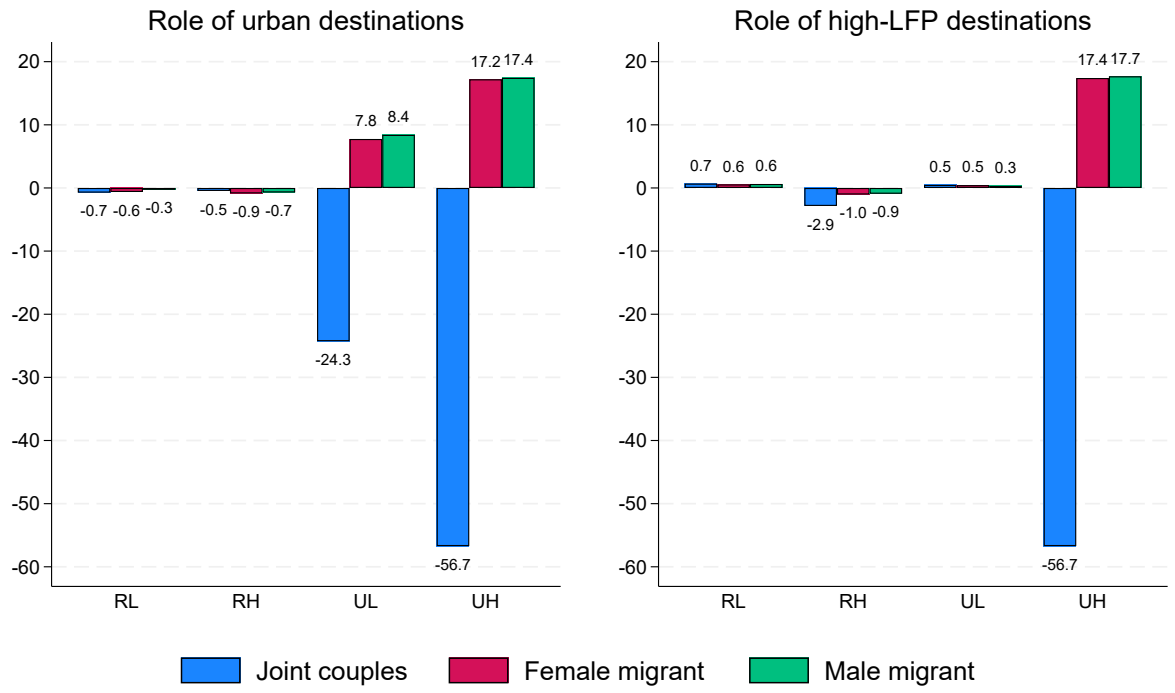


Figure 1.13: Effect of destination characteristics on joint and mixed marriages by location

Notes: Changes from the baseline are calculated as percentage point differences in the share of a type of couple in each location. The graph on the left relates to equation 1.9 and the one on the right to equation 1.10. Locations are either rural (R) or urban (U) and have either low (L) or high (H) labor force participation rate.

driven by rural marriage migrants.

Figure 1.15 presents results at the origin for the case where joint migration to destinations with high labor force participation is restricted. For rural origins, those with higher labor force participation rate see similar changes in native and joint migrants, as well as male and female marriage migrants. However, there is limited effect on couples from rural places with lower labor force participation. This is likely because they primarily choose urban locations that have a lower level of labor force as well, given that disutility of migration between these locations is low (as shown in Table 1.5). Therefore, their preferred urban destinations are no longer restricted and they do not need to adjust their marriage decision.

Overall, the results suggest that joint migration restrictions at the destination can have large effects on the marriage market. While some couples may choose other destinations, especially urban entry restrictions lead not only to couples remaining at the origin, but also to more mixed couples forming at the destination. This implies that marriage migration can be an alternative channel to realize migration intentions when joint migration is impossible.

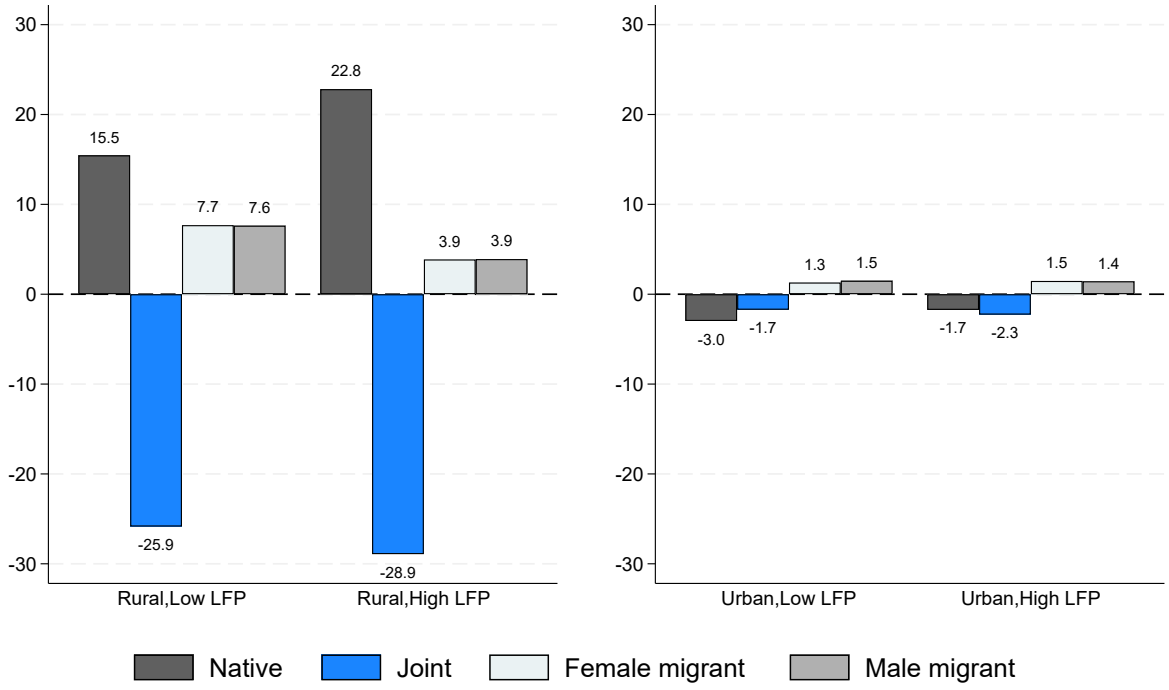


Figure 1.14: Effect of urban destination on marriage choices by origin

Notes: The graph depicts the difference of the share of each category in the counterfactual to their baseline shares. The calculations are based on the counterfactual scenario where joint migration to urban destinations is restricted.

1.7 Conclusion

This study examines how marriage and migration decisions interact with each other in a setting with large internal migration flows and a high prevalence of marriage. In particular, I provide evidence on the role of matching at origin, and how location characteristics enter into the decisions to migrate jointly or independently (for marriage). Thereby, this chapter contributes to our understanding of how migration preferences are integrated into the decision where and whom to marry. By allowing a trade-off between joint migration and marriage migration, it produces novel insights into the preferences for each migration channel under certain constraints.

The analysis is built on a structural matching model with the options to marry and stay at home, migrate together, or marry someone in a different market. The estimation results suggest that matching at home provides a high value for couples, and that couples migrating together face utility losses compared to those staying at home. The type of destination influences the utility level: couples moving to urban areas are better off, and synergies in the level of labor market activity increase marriage utility.

The counterfactual analysis further shows that influencing the utility of joint migrants can affect marriage choices at the origin and at the destination. When joint migration is less costly, less men and women decide to get married at the destination and instead find a spouse at home to migrate together. This affects primarily rural origins and urban destinations. On the other hand, restricting joint access, especially to urban areas, increases mixed marriages. These results suggest that costly migration

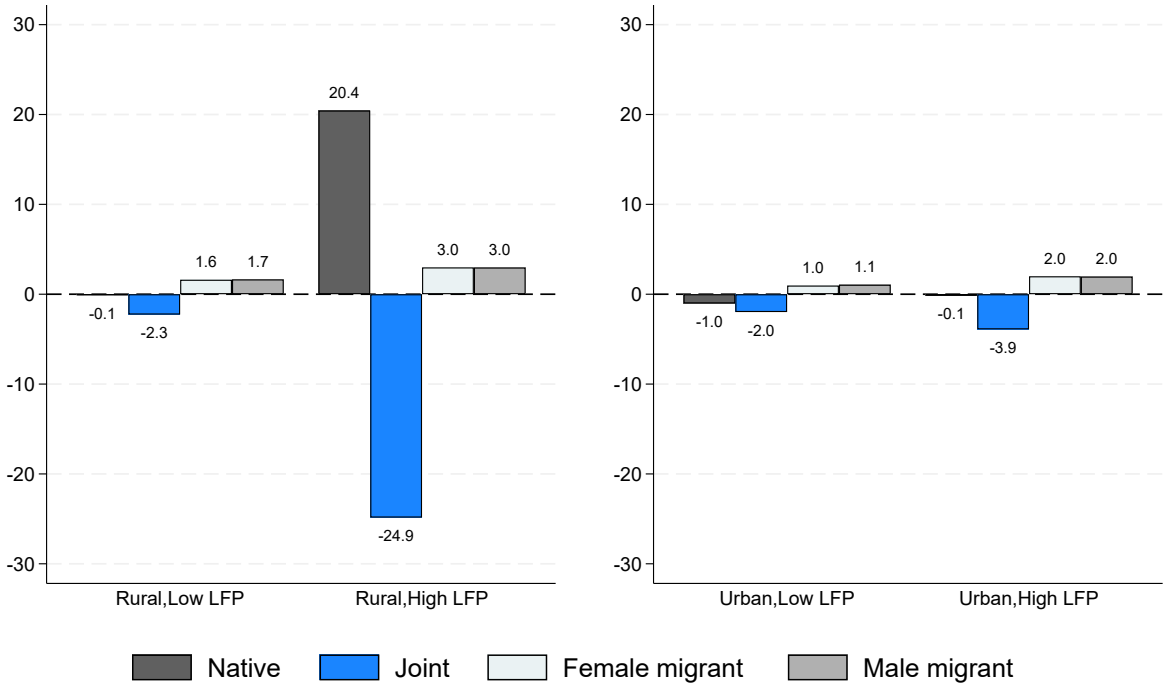


Figure 1.15: Effect of high labor force participation on marriage choices by origin

Notes: The graph depicts the difference of the share of each category in the counterfactual to their baseline shares. The calculations are based on the counterfactual scenario where joint migration to destinations with high labor force participation is restricted.

for couples can affect both men's and women's decisions to marry someone based on more favorable location characteristics.

The findings highlight potentially unintended effects of different migration and development policies. In particular, improving urban amenities and living conditions that favor families may disproportionately incentivize couple migration. On the other hand, policies aimed at curbing or navigating the relocation of entire family units may lead to more independent migration and mixed marriages in urban areas. Further research may test how these decisions affect overall population dynamics and household welfare.

Appendix

1.A Figures

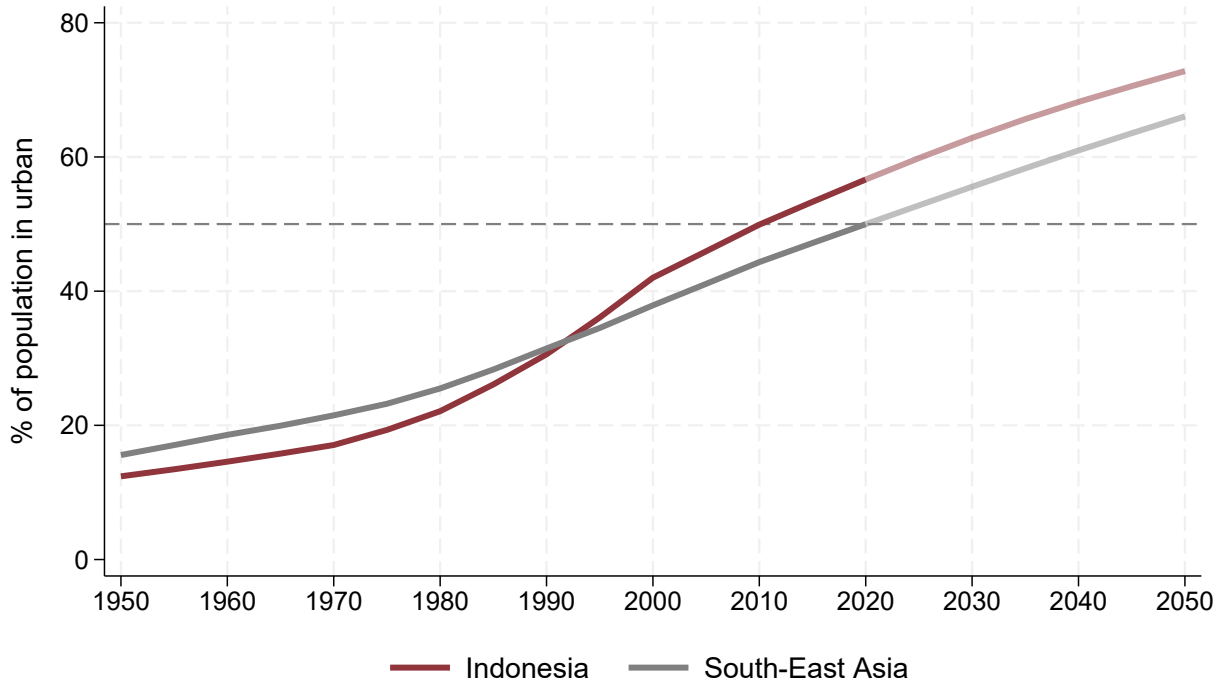


Figure 1.A.1: Urbanization trends in Indonesia and South-East Asia

Notes: Data from UNDP World Urbanization Prospects 2018. The dashed line indicates the 50% threshold of urban population. Indonesia surpassed the regional average for urbanization in the early 1990s, reaching an urban population of over 50% by 2010.

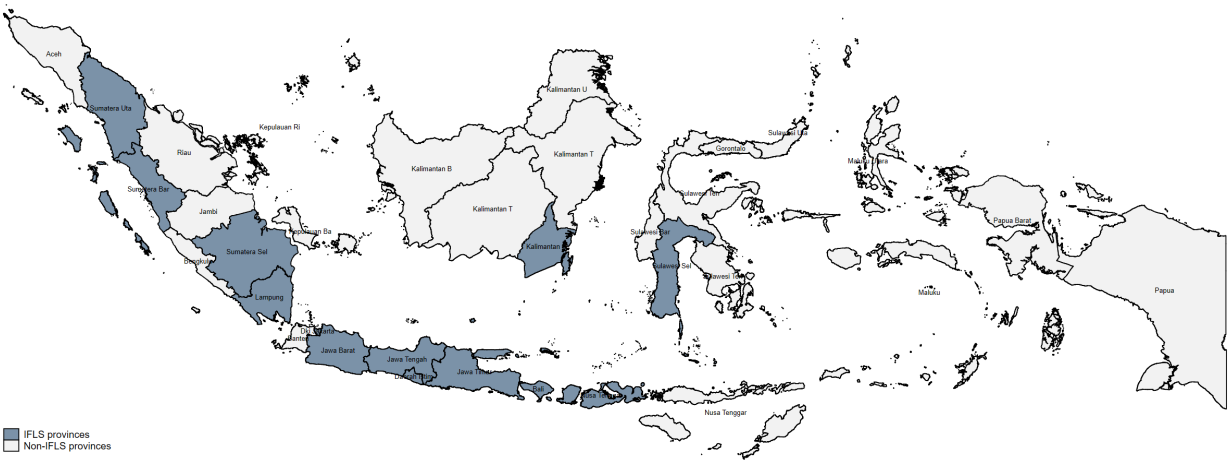


Figure 1.A.2: Selected provinces in the Indonesia Family Life Survey

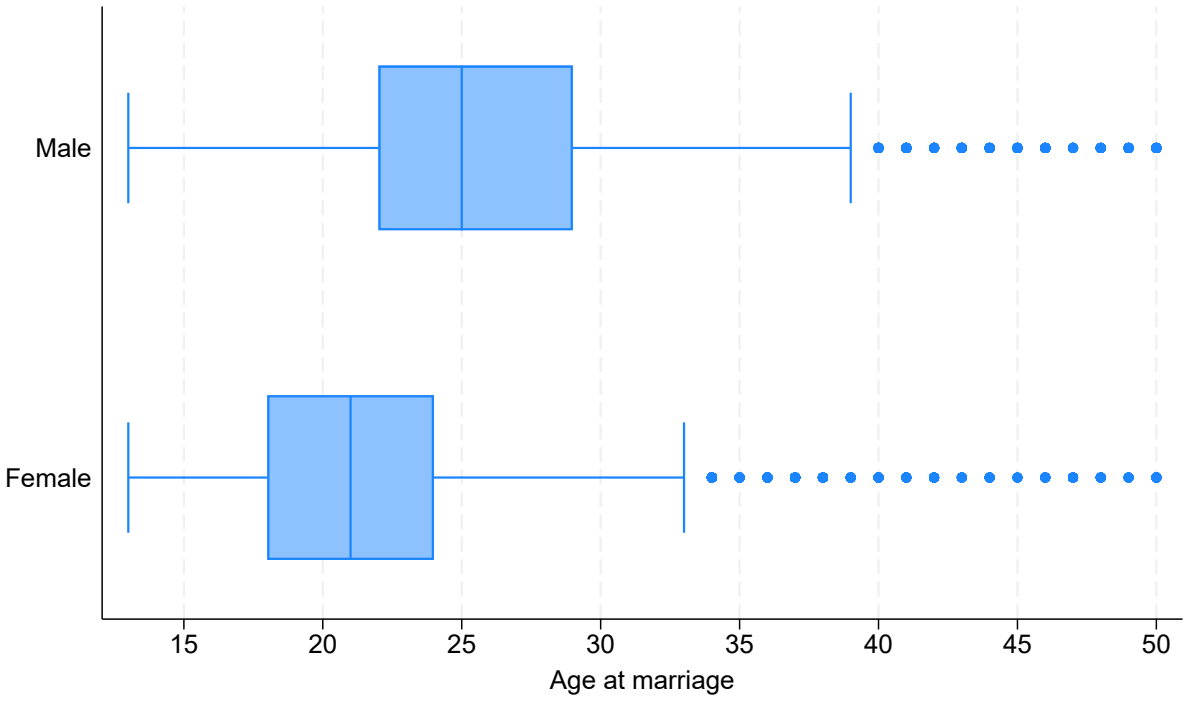
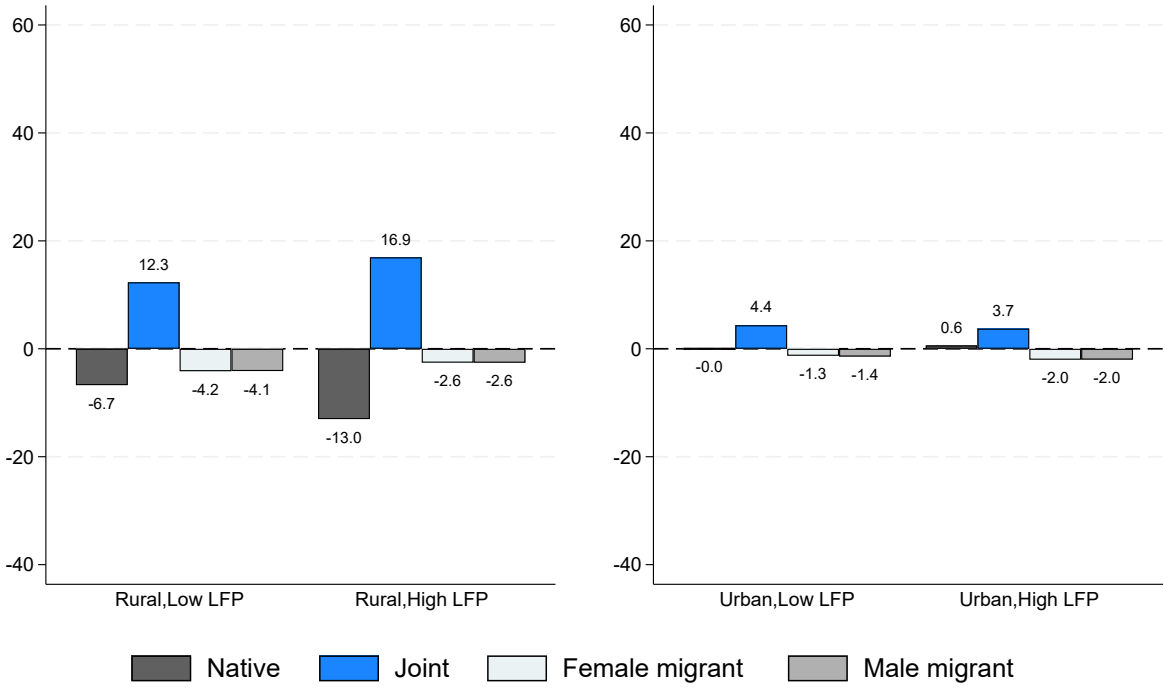
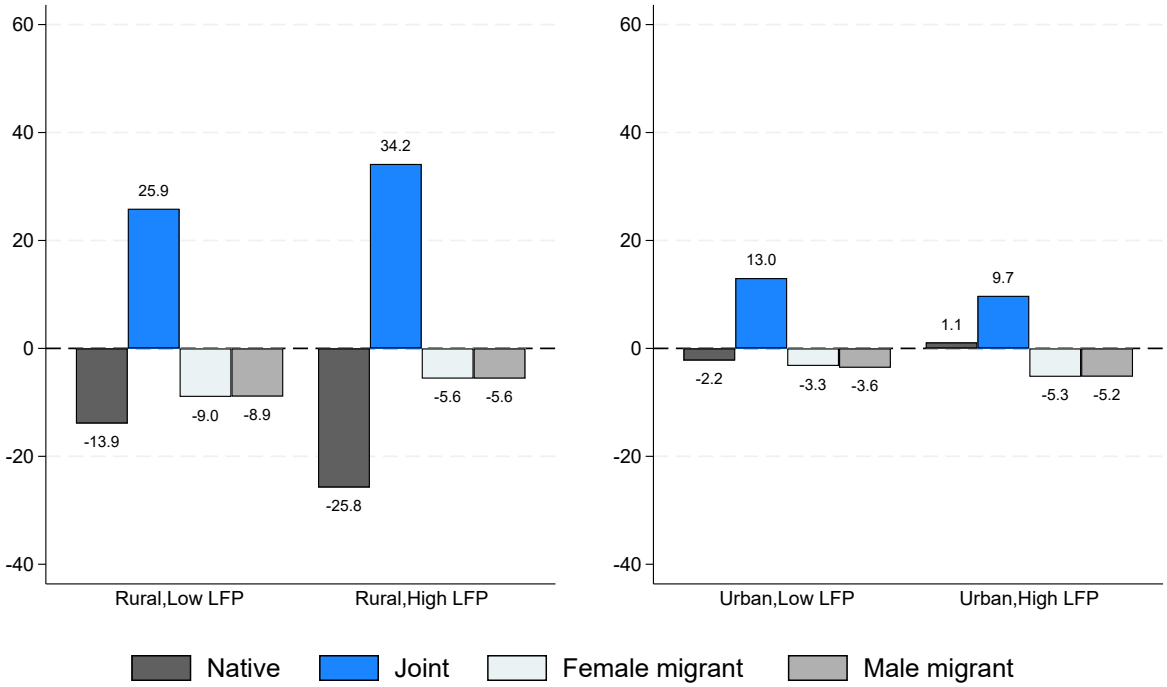


Figure 1.A.3: Distribution of age at marriage for men and women

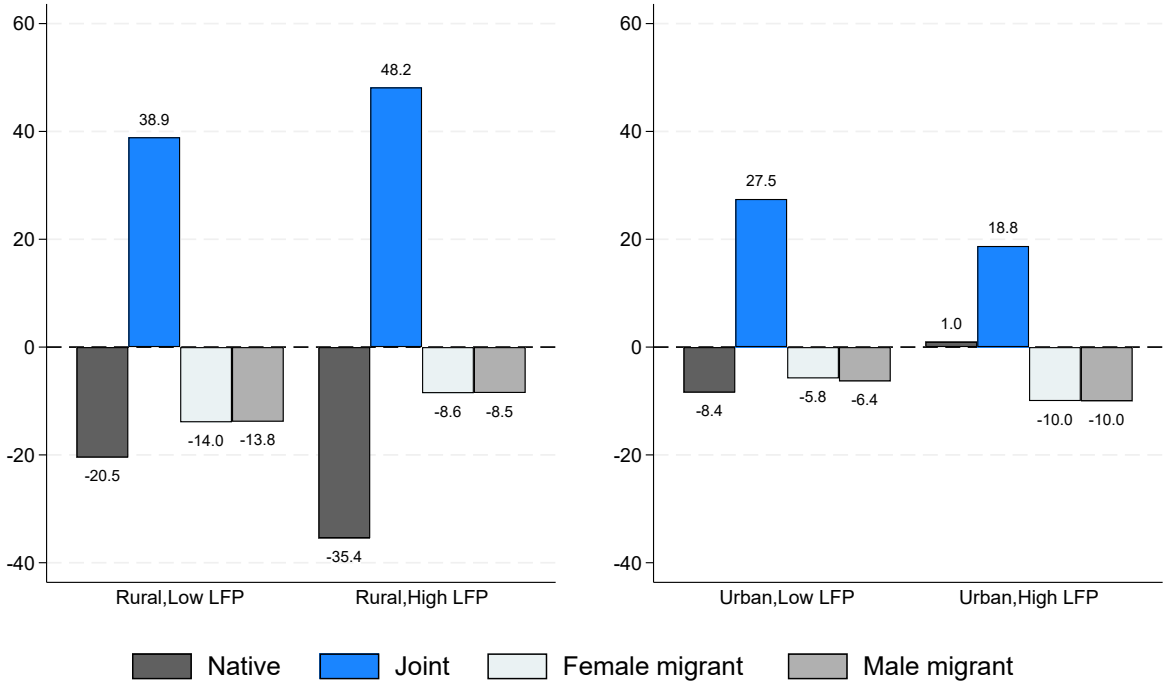


(a) Changes at 25% of native utility added



(b) Changes at 50% of native utility added

Figure 1.A.4: Effect of partially reduced joint migrant disutility on marriage choices by origin



(a) Changes at 76% of native utility added

Figure 1.A.4: Effect of partially reduced joint migrant disutility on marriage choices by origin

Chapter 2

Urbanization and Interethnic Marriages in Sub-Saharan Africa

2.1 Introduction

Sub-Saharan Africa is undergoing rapid urbanization, with the potential to transform social interactions and to affect identity formation in a multi-ethnic context. Urban spaces intersect with ethnic diversity and the composition of ethnic groups, thereby offering potential for increased interactions, including in the marriage market. Local marriage market conditions have been hypothesized to play a key role in the formation of interethnic unions (Kalmijn, 1998). Intermarriage is often used as an indicator of social cohesion, attenuating intergroup conflict and improving economic outcomes (Basu, 2015; Demarest & Haer, 2022; Pengl et al., 2022). Currently, about 20% of marriages in Sub-Saharan Africa are interethnic and more people might form unions with individuals from different ethnic backgrounds as the region urbanizes (Bandyopadhyay & Green, 2021; Crespin-Boucaud, 2020).

This chapter explores the links between urbanization and ethnic assortative mating in Sub-Saharan Africa, highlighting how urban environments facilitate cross-ethnic unions and contribute to broader societal changes. We use Demographic and Health Surveys (DHS) collected in multiple waves between 1990 and 2023 across 25 Sub-Saharan African countries to investigate the links between urbanization and interethnic marriage. These repeated cross-sectional data are geocoded and provide ethnicity information on couples formed between 1975 and 2023. While urbanization is measured contemporaneously in the DHS and in most existing studies (Bandyopadhyay & Green, 2021; Crespin-Boucaud, 2020), we link DHS data with the Global Human Settlement Layers (GHS) data to measure the urban status of places at the time of union formation. The GHS is geocoded at pixel level and provides a harmonized classification of the degree of urbanization across countries from 1975 to 2015.

Descriptively, we find two sets of results. First, we show interethnic marriages increase over marriage cohorts in all regions of Sub-Saharan Africa. While about 15% of our sample marry people of a different ethnic group in 1975, the ratio reaches 25% of couples formed after 2010. The highest rate of interethnic marriage, about 30% in 2010, is observed in Western Africa. Second, there are large differences in the intermarriage rate in urban and rural areas, with the urban share of interethnic couples almost twice as high as for rural couples. Over marriage cohort, the rate of interethnic marriage is constantly higher in urban areas compared to rural areas.

Results from our linear probability models suggest that, on average, living in an urban area is associated with an increase in the probability to marry a person from a different ethnic group of 7-8 percentage points. Using a more granular definition of urbanization, we then show that this probability increases gradually with the level of urbanization. This exercise shows that the difference in the probability to be in an interethnic union between very low-density rural areas and urban centers grows to 13 percentage points.

While these results highlight the persistent rural-urban gap in interethnic marriage, the reduced-form analysis does not provide any explanations on how urban spaces influence intermarriage rates. Theory suggests that diverse urban settings may affect the group sizes and exposure to outgroups in the potential pool of spouses, as well as lead to changes in preferences and norms regarding marriage. Indeed, our data shows that urban clusters are more diverse than rural clusters by around 10 percentage points. With respect to exposure and proximity to other ethnic groups, differences in the distribution of groups across rural and urban areas can affect the probability at which two groups meet. On the other hand, marital institutions may play different roles

in rural and urban contexts, leading to changes in preferences for marriage and traits of spouses. To disentangle the role of group sizes and distribution from preferences, we use a two-sided matching model. This model allows us to estimate preferences for endogamous (within ethnicity) or exogamous (interethnic) marriages, while holding constant the distribution of ethnic groups in rural and urban markets. We further introduce education as a socioeconomic trait that may be traded off with ethnicity when finding a spouse.

This exercise shows that, while matching on the education level is of similar relevance in rural and urban markets, matching on ethnicity receives greater weight in rural areas. This suggests that ethnic assortative mating becomes relatively less important with respect to education in urban areas. We further employ the model to predict how marriages would form differently in urban areas if preferences or distributions were as they are observed in rural areas. Thereby, we can quantify the role of these distinct channels in explaining the gap between intermarriage rates in urban and rural areas.

In particular, we test the relevance of three explanations: 1) differences in group distributions; 2) differences in the total marriage surplus, i.e., utility from marriage and preferences within marriages; 3) differences in the preferences for ethnic traits when utility from marriage is held constant. We find that a large difference in interethnic marriages between rural and urban areas can be explained by stronger preferences for endogamous, i.e., same-ethnicity, marriage in rural areas. A potential explanation is that urban dwellers tend to be less traditional in their marital norms and that more intergroup contact in urban settings erodes outgroup bias. Therefore, cities can enhance social inclusion and enable primary group relations, including intermarriage.

Related Literature These findings contribute to several strands of the literature. First, this chapter provides evidence on how development processes can affect interethnic marriage as a pertinent measure of social cohesion. In Africa, social cohesion measures such as trust, political stability and intermarriage have been linked to early contributors including the slave trade (Nunn & Wantchekon, 2011), ethnic fragmentation (Easterly & Levine, 1997) or export agriculture and print technologies (Pengl et al., 2022). Bandyopadhyay and Green (2021) explore the determinants of interethnic marriage in Sub-Saharan Africa at the country level, finding positive associations with indicators of modernization, such as urbanization and education. At the local level, there has been mixed evidence on how ethnic diversity affects intergroup contact or conflict. For example, Bertinelli et al. (2025) find a positive effect of increased polarization on conflict. Other studies focusing on the urban context highlight the role of points of interaction and the design of urban spaces themselves (Wessel, 2009). To our knowledge, this study is the first quantitative exploration of the role of local conditions, in particular the degree of urbanization, on intergroup contact in the form of intermarriage in Sub-Saharan Africa.

Second, we show the relative importance of matching on education or ethnicity for marriages in rural and urban contexts. While other studies have either focused on the evolution of education assortative mating (e.g., see Eika et al. (2019) for Europe and Pesando (2021) for Sub-Saharan Africa) or interethnic and interfaith marriages (Crespin-Boucaud, 2020), we bring together the two as complementary predictors of marriage. Bandyopadhyay and Green (2021) show that education homogamy (i.e., having the same education level) is negatively associated with interethnic marriage, a finding we confirm in our empirical analysis. Beyond this, we show that, when holding distributions of both education and ethnic groups constant, ethnic preferences,

compared to education, play a larger role in rural marriages than for those in urban areas.

Finally, the last part of this chapter focuses on disentangling the effect of preferences and group exposure on interethnic marriage. With this, we align with a developing literature on the role of meeting opportunities on assortative mating (Belot & Francesconi, 2013; Ciscato, 2024; Jaffe & Weber, 2019; Kalmijn & Flap, 2001; Wen & Mao, 2025). Closest to our approach is Goldman et al. (2025), investigating the role of segregation by race and class on intermarriage in the U.S.. Our context differs from theirs in several ways. First, while they consider interracial marriages between whites and blacks, interethnic marriages in Sub-Saharan Africa may constitute a different type of choice set, defined by linguistic as well as cultural distance. Second, as urbanization plays a major role in shaping the contemporary society in Africa, we focus on differences between urban and rural localities. Third, we contrast the role of exposure to other groups with the role of changing preferences for intermarriage. Interestingly, their results align with our findings in the sense that exposure has only minimal effects on intermarriage.

The rest of this chapter is structured as follows. Section 2.2 provides a conceptual framework from existing literature outlining the potential effects of urbanization on interethnic marriage formation. Section 2.3 describes the data sources and sample, as well as trends in interethnic marriages across Sub-Saharan Africa. Section 2.4 then outlines the empirical strategy and its results of the effects of urbanization on interethnic marriage. In section 2.5 we further explore the potential channels of this effect using a structural model. Section 2.6 finally provides some concluding remarks.

2.2 Conceptual framework: The role of urbanization for interethnic marriages

Urbanization reshapes African marriage markets by altering who meets whom, who can marry whom, and at what social cost. Urban areas may provide greater exposure to media, education and diverse neighborhoods, fostering changes in social norms and attitudes towards interethnic unions. They tend to be more economically developed with better access to labor market opportunities, thereby lowering the reliance on kinship ties for economic advantage. As outlined in Kalmijn (1998), intermarriage can be influenced by three main social forces: individual preferences, group norms, and marriage market constraints. These principles are in line with the economic theory of union formation brought forward by Becker (1973, 1974), which places spouses as utility maximizers and the constraints of the market supplies at the forefront of marriage realization. In these models, ethnic homogamy may bring benefits due to shared language, culture, or social networks. Several interlinked mechanisms related to these considerations may affect interethnic marriages in cities.

(1) Intergroup contact and exposure. Urban living arguably increases the points of interaction between different groups, for example through the education system, neighborhoods or workplaces (Wessel, 2009). This type of intergroup contact may enhance perceptions of out-groups (Allport, 1954) as well as erode in-group identification, thereby fostering intermarriage (Kalmijn, 1998). While on the national level some studies have found increased conflict potential due to higher ethnic fractionalization, other evidence suggests that the geographic level of observation is key, with community-level

diversity reversing this relationship (Fielding, 2024). Several (quasi-)experimental studies have confirmed the importance of contact for reducing prejudice and strengthening ties between ethnic or religious groups (Bazzi et al., 2019; Bursztyn et al., 2024; Lowe, 2021). Importantly, this channel implies that groups are not fully segregated in their urban neighborhoods (Robinson, 2020).

(2) Population size and marriage market supplies. Cities and towns are distinguished by higher population density compared to rural areas. Assuming that the number of men and women increase proportionally, the probability of meeting a potential spouse from the same ethnic group decreases mechanically with higher diversity levels (Kalmijn, 1998). This implies that even without changing attitudes towards other groups, urban marriage markets may exhibit different intermarriage levels. However, if preferences for marrying within one's group are stronger than the prioritization of marriage, this may also affect the incidence of marriage overall rather than intermarriage.

(2) Cultural and socioeconomic preferences. Preferences for marrying within the same ethnic group may stem from traditional practices and the importance of kinship ties, but also cultural similarity and mutual understanding (Kalmijn, 1998). In more developed regions, these preferences may become less relevant compared to economic markers. For example, changing educational attainment for men and women may come with stronger probability of assortative mating in education (Blossfeld & Timm, 2003). Empirical evidence shows increasing education homogamy in Sub-Saharan Africa, in line with development and urbanization processes in these countries (Pesando, 2021). Urban areas offer the potential, and possibly increase the necessity, of higher labor market attachment of women compared to rural areas. This brings forth the hypothesis that women's higher education becomes more important in marriage (Blossfeld, 2009), thereby fostering educational matching at the expense of ethnic matching.

(3) Norms and individual autonomy. Traditional marriages in Sub-Saharan Africa are characterized by extensive family involvement in the selection of spouses. Exogamous marriages may be sanctioned by the family or wider community (Kalmijn, 1998). This firm involvement is thought to be stronger in rural areas, with urban areas characterized by cultural openness and more modern lifestyles (Hamon & Ingoldsby, 2003). Luke and Munshi (2006) find increased agency in the marriage decision by male urban migrants in Kenya, particularly for highly skilled individuals that are less reliant on kinship networks to find employment through the spouse's family. Rather than changing who to marry, the study finds negative selection into marriage in terms of labor market potential.

Empirically, we would expect higher intermarriage rates in urban areas if it indeed affects preferences, intergroup contact through increased diversity, and less reliance on kinship structures. On the other hand, if cities were fully segregated and economic incentives for in-group marriage remains strong, we would not expect any effect. We test the relevance of urban dwelling on the propensity to marry across ethnic lines in our empirical section. To further investigate which channels prevail, we turn to a structural approach in section 2.5.

2.3 Data and descriptive evidence

2.3.1 DHS

The primary data source are the DHS surveys from Sub-Saharan African countries. The DHS are nationally representative household surveys conducted in low- and middle-income countries to collect data on various demographic, health, and population metrics. Initiated in 1984, DHS data are collected through interviews with women, men, and other households members. The structure of the surveys is standardized across countries and years so that the data are comparable across both dimensions. We use data on couples from Sub-Saharan countries where ethnicity information is available. Our definition of intermarriage is based on the self-reported ethnic identity collected from the wife and the husband. The groups are typically based on language-related identification, except in cases where race is more salient, in particular South Africa and Zimbabwe.

Table 2.3.1 gives an overview of the 25 countries used in our study, the number of survey years and the number of observed couples by country, as well as the share of the sample living in urban areas and in an interethnic marriage. Our data spans across 32 years and includes a total of 202,922 couples. The average share of urban (married) population is 30.3%, with the highest urbanization observed in Gabon and South Africa, and the lowest in Chad and Malawi. The interethnic marriage share varies from 4.1% and 7.8% in South Africa and DRC, to 33.2% and 36.5% in Kenya and Uganda.

We further use cluster identifiers, i.e., the primary sampling unit in which households are observed in each survey round. Survey clusters are georeferenced, which allows us to link the location of couples to information on urbanization as outlined below. On average, countries are divided into 470 survey clusters, with an average of 6 households observed in a cluster.

2.3.2 GHS-SMOD

An important limitation of the DHS data is that the urban-rural divide in DHS is defined administratively and is measured at the time of the survey. As a consequence we do not have the information at the time of marriage of most couples. To overcome this problem[16], we use georeferenced survey clusters to link DHS data with the Global Human Settlement Layers (GHS) data. The GHS provides a harmonized classification of the degree of urbanization across countries from 1975 to 2015, collected in 5 year intervals. The area is classified into various urban or rural clusters based on population density and size as well as built-up area density with a resolution of $1km^2$ per cell. The GHS classifies the grid cells into urban center, dense urban cluster, semi-dense urban cluster, suburban or peri-urban grid cells, rural clusters, low density rural grid cells, and very low density rural grid cells (Florczyk et al., 2019). We link the DHS geo-coded survey clusters to the GHS grid cell by year of marriage for each couple in the cluster.[17] This exercise presents several advantages. First, it offers a harmonized,

[16] Under the assumption that the survey cluster in which we observe the couple is the same as the one they got married in.

[17] Given that we have updated urbanization data every five years since 1975, we link the marriage year based on the following rule: Assign the classification of the closest GHS year in terms of absolute difference to the marriage year, and assign the value of 1975 for all marriage years prior to 1975.

Table 2.3.1: Sample summary

Country	Years	Obs.	Urban	Intermarriage
Central Africa				
Cameroon	5	8892	0.421	0.245
Central African Republic	2	939	0.334	0.240
Chad	2	2955	0.184	0.156
DRC	3	6833	0.303	0.078
Gabon	4	3408	0.611	0.435
Eastern Africa				
Ethiopia	4	17125	0.209	0.123
Kenya	5	13402	0.348	0.183
Malawi	6	11095	0.149	0.332
Mozambique	1	2141	0.362	0.265
Uganda	2	3517	0.183	0.365
Zambia	3	10283	0.382	0.461
Southern Africa				
South Africa	1	663	0.617	0.041
Western Africa				
Benin	6	9778	0.362	0.165
Burkina Faso	7	14885	0.241	0.122
Côte d'Ivoire	4	6768	0.370	0.279
Gambia	2	1897	0.467	0.292
Ghana	9	9901	0.358	0.217
Guinea	3	5833	0.271	0.147
Liberia	1	1844	0.316	0.274
Mali	7	12376	0.250	0.303
Niger	2	4030	0.278	0.155
Nigeria	3	25389	0.322	0.131
Senegal	12	15919	0.307	0.211
Sierra Leone	3	8912	0.295	0.207
Togo	3	4137	0.247	0.123
Total	32	202922	0.303	0.210

Notes: Sample of observed couples in Sub-Saharan Africa by country. The column "urban" provides the share of the couples in urban locations, and 'intermarriage" the share of the couples with different ethnicity.

globally comparable method to delineate urban, rural, and intermediate areas. Cross-country analyses such as ours thereby benefit from not relying on varying national definitions. Second, the classification into different rural and urban categories allows for a more refined analysis of the degree of urbanization. Third, and most importantly for our empirical strategy, we are able to capture the level of urbanization at the time of marriage, as opposed to using contemporaneous measures of urban location. This also allows us to exploit variation over marriage timing within the same cluster.

2.3.3 Interethnic marriages across countries and over time

As shown in Figure 2.3.1, there is a positive association between the share of the urban population and intermarriage for the countries in our sample. Both the urbanization and the intermarriage rate vary widely in our sample, ranging from over 45% intermarried couples in Zambia to less than 10% in Cote d'Ivoire, as well as around 15% urban population in Malawi to over 60% in Gabon and South Africa.[18]

Figure 2.3.1: Country-level correlations between urban population and intermarriage share

From Figure 2.3.2, we observe a general increase in the rate of interethnic marriages across Sub-Saharan African countries in our sample from the 1980s to the 2010s. A particularly striking example is Kenya, where the proportion of interethnic marriages in the 2010s is more than four times higher than in the 1980s. In the 1980s, the countries with the highest proportions of interethnic marriages were Gabon (33.46%), Zambia (37.26%), and Uganda (33.05%), whereas the lowest proportions were observed in Kenya (7.94%), the Democratic Republic of the Congo (6.51%), and South Africa (2.85%).

[18] South Africa presents itself as an outlier with a high urbanization rate and low intermarriage. This corresponds to its exceptional status in the region in terms of development as well as its unique history of racial segregation.

By the 2010s, Zambia (49.75%), Gabon (48.73%), and Uganda (37.64%) exhibited the highest rates of interethnic marriages, while the Democratic Republic of the Congo (9.36%), Namibia (8.72%), and South Africa (4.77%) had the lowest. Notably, despite having a high proportion of marriages in urban areas, South Africa consistently displays a low proportion of interethnic marriages.

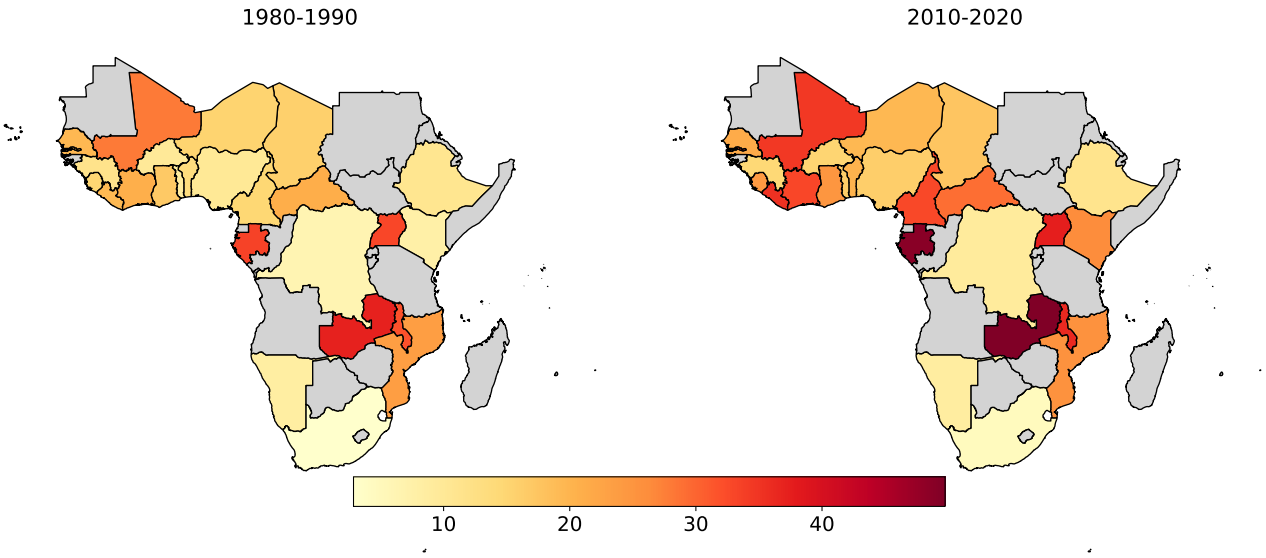


Figure 2.3.2: Average percent of interethnic marriages in Sub-Saharan Africa.

Therefore, most Sub-Saharan African countries in our analysis show an increase in both the proportion of marriages occurring in urban areas and the proportion of interethnic marriages between the 1980s and the 2010s. Figure 2.3.3 compares the aggregate intermarriage and urbanization rates in the region over marriage cohort. We observe a steady increase in intermarriage over time from around 15% in the period before 1975 to about 25% for marriages in 2010 and after. This upward trend follows closely the trajectory of the urban population share over the same time.

Figure 2.3.4 dissects the trend in intermarriage by urban and rural classification from the GHS-SMOD data. While marriages in all types of locations exhibit an upward sloping share of interethnic couples, shares in urban areas are consistently higher and, particularly for dense and semi-dense urban clusters, exhibit a higher growth rate.

2.3.4 Pooled sample statistics

The pooled analysis sample, comprised of married or co-habitating couples, is described in Table 2.3.2, with the statistics of the total sample as well as split by urban and rural residence.

To get an overview of how diversity differs across rural and urban locations, we first build a measure of local diversity using the fractionalization measure proposed by Alesina et al., 2003. The general index is calculated as

$$FRACT_j = 1 - \sum_{i=1}^{N} s_{ij}^2$$

where s_{ij} is the share of group i in locality j. While the index serves as a country-level fractionalization measure, we replicate the exercise on the survey cluster level to

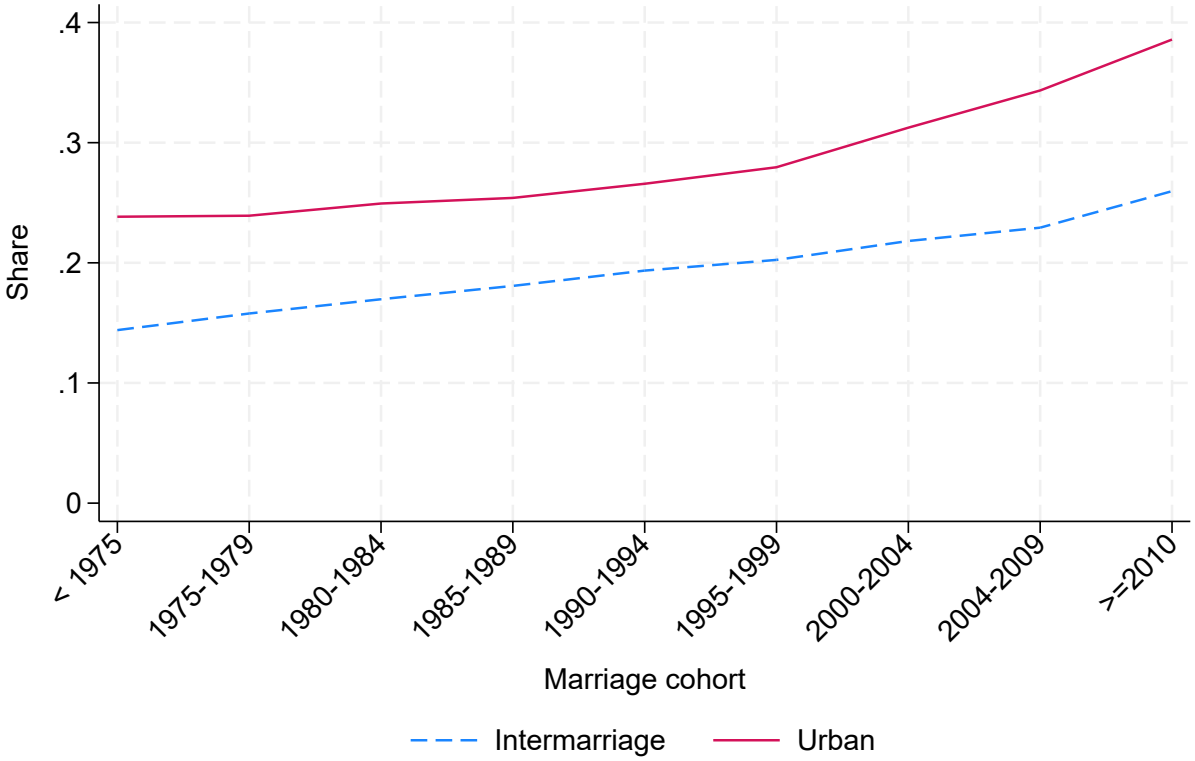


Figure 2.3.3: Urbanization and Intermarriage over marriage cohorts

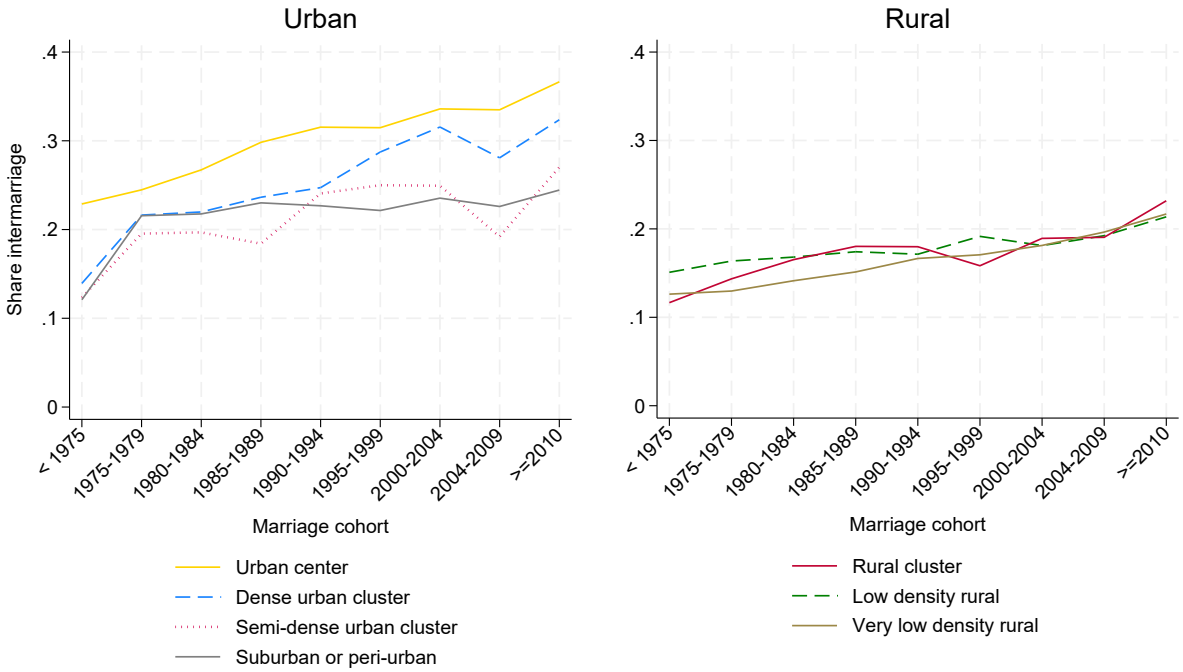


Figure 2.3.4: Intermarriage by urban classification over marriage cohorts

provide an understanding of local exposure to different ethnic groups. A higher index (bounded between 0 and 1) implies higher fractionalization or, in other words, diversity.

Comparing the average diversity of urban and rural clusters, urban clusters are more diverse by about 10 percentage points.

Table 2.3.2: Summary statistics (Pooled DHS couple data)

	Total		Rural		Urban	
	Mean	SD	Mean	SD	Mean	SD
Cluster						
Diversity	0.470	0.276	0.364	0.276	0.457	0.282
Couple						
Interethnic	0.210	0.407	0.168	0.374	0.306	0.461
Same Education	0.608	0.488	0.642	0.479	0.529	0.499
Wife						
No education	0.516	0.500	0.604	0.489	0.314	0.464
Primary education	0.278	0.448	0.279	0.449	0.274	0.446
Secondary or higher	0.206	0.405	0.117	0.321	0.411	0.492
Age at marriage	17.879	4.205	17.307	3.850	19.194	4.665
Current age	30.613	8.239	30.430	8.360	31.034	7.939
Husband						
No education	0.403	0.491	0.481	0.500	0.224	0.417
Primary education	0.285	0.452	0.307	0.461	0.236	0.425
Secondary or higher	0.312	0.463	0.212	0.409	0.540	0.498
Age at marriage	24.392	5.603	23.765	5.381	25.831	5.832
Current age	38.728	9.557	38.621	9.758	38.975	9.071

Notes: The sample consists of married couples with women aged 15-49 and men aged 15 to 97. This discrepancy in age groups is due to the fact that the primary population of interest are women aged 15-49, and husbands of any age are linked to the interviewed woman. Diversity measure calculated following Alesina et al., 2003 at cluster level.

Similarly, we observe a larger share of interethnic couples in urban (30.6%) compared to rural (16.8%) locations. Conversely, more rural couples have the same education than urban couples. This may reflect different preferences as well as greater similarity in education attainment of men and women. Indeed, we see that, compared to the urban sample, rural husbands and wives have a higher concentration in the "no education" category. Generally, husbands are more likely to have secondary or higher education than their wives. Yet, female education is more evenly distributed in urban areas. The average age gap in marriage is around 6 years, with the average marriage age for women at 17.5 and for men at 23.5. Both men and women tend to get married on average around 1.5 years later in urban compared to rural settings. Lastly, the current ages of our sampled couples are comparable between rural and urban areas, with an average age of 30.6 for women and 38.7 for men.

2.4 Estimating the rural-urban gap in intermarriage

We now analyze how urban location affects interethnic marriages. For this, we use granular urbanization data over time from the GHS-SMOD dataset to link urban status to a couple's cluster at the time of their marriage. Thereby, we exploit within-cluster variation over time to study the effect of urbaniztion on the probability of interethnic marriage.

2.4.1 Empirical strategy

The probability of the marriage being between spouses with different ethnicities is estimated by the linear probability model

$$Inter_{icmt} = \alpha_c + \beta U_{cm} + C_i'\gamma + \delta_m + \lambda_t + \varepsilon_{icmt} \tag{2.1}$$

where $Inter_{icmt}$ is an indicator variable that equals 1 if the marriage is interethnic for couple i in cluster c married in cohort m and surveyed in year t, and 0 otherwise. The coefficent β measures the effect of the variable of interest, U_{cm}, which indicates if cluster c was defined as urban during marriage cohort m. This is either a binary variable where $U_{cm} = 1$ if the cluster is classified as urban, and 0 otherwise; or a categorical variable of the urbanization status based on the classification in GHS-SMOD (low density rural, rural, semi-dense urban, suburban, dense urban, or urban center, with very low density rural as the excluded category). This allows us to compare the crude rural-urban divide with a more refined measure of urbanization that takes into account the gradual nature of urbanization.

The vector C_i' collects characteristics of the couple that might bias the estimated effect of urban location on the probability of interethnic marriage. These are the ages and the education levels of the husband and the wife, as well as an indicator for having the same education level.[19] We add cluster fixed effect α_c[20], marriage cohort fixed effect δ_m, and survey year fixed effect λ_t.

2.4.2 Results

We present the main results from our analysis of the effect of urban location at time of marriage on the probability of intermarriage in Table 2.4.1. Columns (1) and (2) present estimates of equation 2.1 without cluster level fixed effects (but controlling for country), while columns (3) and (4) contain the main specification results at the cluster level. In panel A, the estimates of the main variable of interest, urban location (as a binary or categorical variable), are presented. Panel B collects the results of the role of education on interethnic marriage.

We find that urban residence at time of marriage has a positive effect on intermarriage in both specifications. In particular, living in an urban location increases the probability of marrying a spouse from a different ethnicity by 7.5 percentage points. The results from the categorical measure of urbanization further shows that this effect is increasing in the intensity of urbanization. For example, compared to a very low

[19] We do not control for diversity level in the cluster as this is likely a mediator of how urban location affects interethnic marriage, and therefore constitutes a bad control (Cinelli et al., 2024).

[20] In an alternative specification, we add country fixed effects instead of cluster fixed effects.

density rural cluster, living in a suburban or peri-urban locality increases the probability of intermarriage by 6.7 percentage points. This effect almost doubles for urban centers, which exhibit the highest population density.

Table 2.4.1: Determinants of interethnic marriage (Linear Probability Model)

	(1)	(2)	(3)	(4)
Panel A: Location				
Urban grid	0.087***		0.075***	
	(0.003)		(0.009)	
Low density rural		-0.002		0.020***
		(0.004)		(0.006)
Rural cluster		-0.001		0.037***
		(0.008)		(0.012)
Semi-dense urban cluster		0.006		0.049***
		(0.010)		(0.015)
Suburban or peri-urban		0.025***		0.067***
		(0.005)		(0.011)
Dense urban cluster		0.079***		0.118***
		(0.006)		(0.015)
Urban center		0.124***		0.136***
		(0.004)		(0.014)
Panel B: Education				
Primary education (husband)	0.003	0.002	0.009***	0.009***
	(0.003)	(0.003)	(0.003)	(0.003)
Secondary or higher (husband)	0.034***	0.030***	0.026***	0.025***
	(0.004)	(0.004)	(0.004)	(0.004)
Primary education (wife)	0.021***	0.020***	0.027***	0.026***
	(0.003)	(0.003)	(0.003)	(0.003)
Secondary or higher (wife)	0.054***	0.048***	0.057***	0.055***
	(0.004)	(0.004)	(0.004)	(0.004)
Same education	-0.018***	-0.018***	-0.010***	-0.010***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	200095	200095	200095	200095
Cluster FE	No	No	Yes	Yes
Countries/Clusters	25	25	32431	32431

Notes: Clustered standard errors in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions with survey year, and cohort fixed effects and controlling for the age of the husband and the wife. Dependent variable is interethnic marriage. Models (1) and (2) run specification 2.1 with country fixed effects, (3) and (4) with cluster fixed effects.

At the cluster level, even smaller increases in density have a statistically significant effect on intermarriage, while the effect becomes statistically significant for the country

level specification only once reaching suburban clusters. This implies that large regional variation within a country increases noise when comparing relatively small differences in the degree of urbanization.

The effects of education on interethnic marriage are explored in Panel B. While the education level seems less important in explaining interethnic marriages than the location, a higher level of education of both the husband and the wife is associated with an increase in the probability of intermarriage. Further, the wife having higher education seems to have a larger effect on this probability than the husband's. A potential explanation for this correlation may be that higher education fosters openness towards other groups, through increased intergroup contact and/or increased empathy. On the contrary, having the same education level is associated with a negative probability of intermarriage. This may imply that education assortative mating and ethnic assortative mating move complementarily. However, we cannot conclude from these results which trait is more important for a marital match. Further, preferences for education and ethnic group may be different in rural and urban environments, and the likelihoods of meeting a spouse with a certain educational level or ethnic background may differ between locations as well. To investigate these channels further, we employ a structural approach to modeling these marital choices in Section 2.5.

2.4.3 Robustness checks

Table 2.4.2 presents robustness checks for 3 modifications to our sample or outcome variable. The checks are run on our main specification 2.1 with cluster-fixed effects and using the 2 different urban classifications.

One important limitation of our data is that, while we observe the urbanization level of the survey cluster at the time of marriage, we do not know whether or not spouses have lived in the same location at that time. This could mean that part of the effect we observe comes from reverse causality, where interethnic couples move to places that are, for example, more open to ethnic mixing, and thereby increase the urbanization level. On the other hand, if couples got married in a more urban place and then moved to a rural residence where we observe them now, the effect may be downward biased. We can partially overcome this limitation as, for some survey rounds and countries, we have information on how many years the respondents have lived in the current location of residence. We use this information to construct a sample of couples that have lived in the same location since at least one year before their marriage.[21]

As shown in columns (1) and (2) in Table 2.4.2, the main effect of urban location on interethnic marriage remains positive. However, the coefficient shrinks by about 3 percentage points. This holds true also for the effects of the categorical variable, with the effect of increasing population density within the rural categories becoming statistically insignificant. While these results generally support our findings, they highlight two potential mechanisms: as the effect is generally lower than in the main specification where migrants are included, this indicates that part of the effect may be driven by in-migration of new ethnic groups. This is in line with the explanation that urbanization is in large part due to migration of different ethnicities, creating more diversity. On the other hand, the fact that the effects remain positive, in particular for more urbanized areas, suggests that other urban environmental factors than the pure population channel are at play.

[21] Repeating this exercise with couples where both spouses have never moved reduces the sample further but yields qualitative similar results.

Another aspect to take into account is that polygamy plays a confounding role, as polygamous couples are more likely to reside in rural areas and may also be more likely to be in traditional, endogamous marriages. A sizable share of men, around 25% of the sample, indicate marriage with more than one wife. While only 15% of men are in polygamous unions in urban areas, this share almost doubles (28%) in rural areas. To rule out polygamy as a confounder, we restrict the sample to monogamous couples only in columns (3) and (4). The results remain stable, both in terms of magnitude and statistical significance.

Lastly, we test the robustness of our approach to two alternative measures of intermarriage, based on ethnicities in the Murdock Map or linguistic distance. Our main analysis is based on the definition of ethnic groups in each country by DHS. The definition of an interethnic couple is, therefore, dependent on how many distinct ethnic groups are considered in the survey. For some countries, these groupings further vary over survey rounds. For example, in Cameroon, the DHS lists 10 ethnic groups in 2011 and 147 in 2018. Instead of taking only one survey round per country (such as Bandyopadhyay and Green (2021)) or recoding ethnic categories based on numerical thresholds (such as Crespin-Boucaud (2020)), we link ethnic groups in the DHS to the time-invariant definition in Murdock (1959). This is a commonly referred to source of ethnic grouping based on cultural, geographical, and economic characteristics. Compared to an average of 50.7 groups per country in the DHS, the Murdock Map classifies an average of 26.2 (Müller-Crepon et al., 2021). To link the DHS with the Murdock ethnic groups, we utilize the LEDA package by Müller-Crepon et al. (2021), which leverages the structure of the Ethnologue language tree to provide a flexible link between any two ethnic groups that are linked to the tree. As columns (5) and (6) in Table 2.4.2 shows, the results are again very similar in magnitude and statistical significance.

A different approach is to take the linguistic distance between ethnic groups as a continuous measure of intermarriage. While the classification of groups in the DHS survey is based on salient features of ethnicity, usually related to language, the grouping still poses a somewhat arbitrary threshold to cultural or linguistic distance. It may be, for example, that even though we observe a higher share of couples from different ethnic groups in urban areas, tastes for marrying someone speaking a language close to the own are not different from rural couples. To test this, we again take data from Müller-Crepon et al. (2021) and re-estimate our model using a continuous outcome variable. The results in columns (7) and (8) suggest a qualitatively similar effect, with a positive sign on the urban variable, and growing magnitude by urban density.

These robustness checks give us confidence that, while endogeneity concerns in the effect of urban location on intermarriage cannot be fully accounted for in the identification strategy, the association is stable and relevant. Yet, these reduced-form effects cannot account for the type of channels that lead to higher intermarriage in urban areas, as outlined in section 2.2. Therefore, the next section is dedicated to disentangling the mechanisms, in particular preferences and the role of diversity.

2.5 Exploring channels: A matching model approach

The above results suggest a positive and economically large rural-urban gap in intermarriage. However, the observed share of marriages between different ethnic groups encompasses not only preferences, but also marriage shares in the population and the

Table 2.4.2: Robustness checks

	Non-movers		Monogamous		Murdock Ethnicity		Linguistic Distance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Urban grid	0.045**		0.082***		0.068***		0.026***	
	(0.018)		(0.010)		(0.009)		(0.004)	
Low density rural		0.003		0.021***		0.019***		0.007**
		(0.013)		(0.008)		(0.006)		(0.003)
Rural cluster		0.010		0.038**		0.030**		0.008
		(0.023)		(0.015)		(0.012)		(0.006)
Semi-dense urban cluster		0.008		0.063***		0.033**		0.012*
		(0.032)		(0.018)		(0.015)		(0.007)
Suburban or peri-urban		0.039*		0.072***		0.056***		0.024***
		(0.023)		(0.012)		(0.011)		(0.005)
Dense urban cluster		0.089**		0.126***		0.101***		0.043***
		(0.035)		(0.018)		(0.015)		(0.007)
Urban center		0.116***		0.144***		0.130***		0.044***
		(0.032)		(0.015)		(0.014)		(0.006)
Observations	52763	52763	150814	150814	160778	160778	169398	169398
Cluster FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	16491	16491	31701	31701	28762	28762	29726	29726

Notes: Clustered standard errors in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All regressions with survey year, and cohort fixed effects and controlling for: age of the husband and the wife, education level of husband and wife, and husband and wife having the same education level. Models (1) and (2) run specification 2.1 only for couples that have never moved, (3) and (4) only for monogamous couples, and (5) and (6) takes the ethnic group based on the Murdock Map.

distribution of different ethnic groups. Therefore, when comparing marriages between rural and urban locations, we cannot conclude from the previous findings if the gap results from differences in preferences or distributions. As outlined in section 2.2, several channels may be at play in explaining differences in intermarriage shares, including differential group exposure, changing preferences for marriage, and for cultural or socioeconomic traits in spouses.

We explore these channels using a structural approach to disentangle the role of assortative mating, i.e., preferences for marrying someone from the same ethnicity, and the role of utility or gains from marrying, from group distributions. In particular, we employ a matching model with transferable utility and idiosyncratic tastes a la Choo and Siow (2006) where men and women can choose a spouse based on their ethnicity and education level. As outlined in section 2.2, education may be seen as an indicator or predictor of socioeconomic status, which could affect the relative importance of ethnic affiliation in marriage. Using data on observed matches and singles, we estimate utility from matching on these spousal traits in rural and urban marriage markets. Thereby, we compare preferences for ethnicity and education in the two markets, holding constant the different supplies of men and women with these attributes. We then use the model to predict how couples would form in the urban market if either preferences or group sizes were equivalent to those in rural markets. By comparing the counterfactual marriage market equilibria with the intermarriage rates under observed preferences, we quantify the role of differences in diversity, preferences and gains from marriage.

2.5.1 Set up

There are two markets in the economy, one rural and one urban. In each market, we observe men and women that can find a spouse within their own market.[22] In the following, we explain the model for a representative market, and all considerations hold true for both urban and rural markets.

We assume a two-sided, one-to-one matching market, i.e., one man can match with one woman to form a union. Men and women on each side of the market are characterized by a vector of attributes, $x \in \{1, ..., X\}$ for men and $y \in \{1, ..., Y\}$ for women. In our application, these attributes contain information about the ethnic identity and education level. Therefore, a man of type x refers to a man who possesses a certain set of observable characteristics that all men of type x share. Similarly, we can define a woman of type y following the same reasoning. For our model, we will consider x and y to be discrete types.

We observe the equilibrium in this market of heterosexual couples where $\mu(x, y), \mu(x, 0)$ and $\mu(0, y)$ are equilibrium outcomes. $\mu(x, y)$ refers to the number of matches between men of type x and women of type y. Therefore, $\mu(x, y)$ will refer only to matches of married population. For singles, $\mu(x, 0)$ refers to the count of men type x, and $\mu(0, y)$ refers to the count of women type y that are single in the market. The set of matches $\{\mu(x, y), \mu(x, 0), \mu(0, y)\}$ must satisfy the feasibility constraints:

$$N(x) = \sum_y \mu(x, y) + \mu(x, 0), \quad \text{and} \quad M(y) = \sum_x \mu(x, y) + \mu(0, y), \tag{2.2}$$

[22] This condition essentially restricts mobility between rural and urban markets and is imposed due to data limitations, as we do not have data on where people have lived before their current residence. Relaxing this restriction would imply that men and women could choose a spouse across markets, thereby affecting the supply of types in each market. This option is topic of Chapter 1.

where $N(x)$ are the number of men and $M(y)$ the number of women of each type in the market.

Men and women in this model are assumed to be utility maximizers and the equilibrium matches we observe are the result of the following utility optimization problems. Man i of type x will choose the type of woman y that maximizes his utility function u_i as given by:

$$\max_{y \in Y_0} u_i(x) = \{\alpha(x,y) + \epsilon_i(y); \quad \epsilon_i(0)\}$$

where $\alpha(x,y) + \epsilon_i(y)$ is the indirect utility from marriage and $\epsilon_i(0)$ is the reserve utility from staying single, with $\alpha(x,0)$ normalized to 0. Symmetrically, a woman j of type y maximizes her utility from marriage v_j by choosing the type of man x that optimizes:

$$\max_{x \in X_0} v_j(y) = \{\gamma(x,y) + \eta_j(x); \quad \eta_j(0)\}$$

The values of $\alpha(x,y)$ and $\gamma(x,y)$ represent the deterministic or direct utility derived from a marriage between a man type x and a woman type y, i.e., they are independent of individual characteristics of man i and woman j. The terms $\epsilon_i(y)$ and $\eta_j(x)$ represent the stochastic part of the utility functions and collect the individual (unobserved) preferences of man i for a type of woman y or of woman j for a type of man x, respectively. Both $\epsilon(y)$ and $\eta(x)$ are i.i.d. type I Extreme Value (Gumbel) random variables, which is a common distributional assumption in discrete choice modeling.

We ensure that the equilibrium matching $\{\mu(x,y), \mu(x,0), \mu(0,y)\}$ is stable by introducing transferable utility dependent on types x and y of a match, $t(x,y)$ (Becker, 1973; Shapley & Shubik, 1971). Without frictions, the transfer cancels out when adding the direct utilities for couple (x,y), allowing us to identify a joint utility from a match between x and y without observing equilibrium transfers:

$$\Phi(x,y) = \alpha(x,y) + t(x,y) + \gamma(x,y) - t(x,y) = \alpha(x,y) + \gamma(x,y) \tag{2.3}$$

Since $\Phi(x,y)$ is the sum of each partner's systematic benefit from marriage, any individual who remains single effectively has zero marital surplus: $\Phi(x,0) = 0$ for single men and $\Phi(0,y) = 0$ for single women. Consequently, $\Phi(x,y)$ captures the total (joint) surplus contributed by both partners' characteristics when they form a couple; in other words, it is a function of the combined characteristics of the husband and wife. By the Gumbel distributional assumptions of idiosyncratic tastes, we get the following non-parametric surplus function (Choo & Siow, 2006):

$$\Phi(x,y) = \ln\left(\frac{\mu^2(x,y)}{\mu(x,0)\mu(0,y)}\right) \tag{2.4}$$

linking the surplus to the observed matches in the market.

2.5.2 Data structure and variables

As mentioned above, we are interested in estimating preferences for matching based on education and ethnicity in urban and rural markets. Therefore, the variables determining a match are education and ethnic group. We denote the first dimension, education, with $Educ_x$ for men and $Educ_y$ for women, which either takes value L (low, not completed primary) or H (primary level or higher), i.e., $Educ_x, Educ_y \in \{L, H\}$.

The second dimension, ethnicity, is denoted by $Ethn_x$ and $Ethn_y$ for men and women respectively, with values G_1 or G_2, i.e., $Ethn_x, Ethn_y \in \{G_1, G_2\}$ as defined below.

To be able to answer the desired questions with our cross-country data, we need to make two adjustments. First, in order to identify reservation utilities and the marriage surplus, we require data on singles by type. For this, we make use of the fact that the DHS not only collects data on all women in the household older than 15, but also a sample of (unmarried) men. We select all households for which this data exists and use information on single men and women as well as couples linked within the household (as before). Second, we need to define an aggregate grouping of ethnicities that allows us to define an "ethnic group" type across all countries. We exploit the fact that countries typically consist of a majority ethnic group in terms of population size and will, therefore, define ethnicity types in terms of belonging to the majority group (or a linguistically close group, G_1) or a non-majority group (G_2). To do this, we first identify the majority group by population size in the DHS sample (counting all married and unmarried individuals). We then employ the LEDA package (Müller-Crepon et al., 2021) to link all groups to the majority group by their linguistic distance.[23] We classify all groups that are either the majority or have a distance to the majority of 0.25 or lower into one group, and the rest into another group.[24] With these adjustments, we aggregate observed men and women, either married/ cohabitating or single, into an urban and a rural market.

Lastly, we select a sample that mimics as closely as possible an equilibrium in the marriage market. To that end, we use a sample of women aged 22 to 46 (the oldest age recorded in the data) and men aged 27 to 52, assuming that the marriage market has cleared in these age groups and those that are observed as single will not get married in the future. Our data confirms that 90% of women (men) were married by age 22 (27). The data also records an average age gap between husbands and wives of around 6 years, thus we restrict the male sample to those aged under 52 years. Further, to follow the 1-to-1 matching condition, i.e., each man and woman can only match with one spouse, we restrict the sample to those couples in monogamous marriages.

Observed Matches

With these data adjustments, we observe around 29,000 men and women in rural areas, and 15,000 in urban areas (Table 2.5.1). In rural areas, 93% of men and 96% of women are married, whereas a lower share of ca. 80% of men and 82% of women are married in urban areas. 60% of rural men and 45% of women are considered having high education (primary level or higher), whereas this share rises to 85% for men and 76% for women in urban areas. Lastly, we observe roughly a 3 percentage point higher share of men and women belonging to a majority (or similar) ethnic group in urban compared to rural areas.

Figure 2.5.1 illustrates the proportion of men and women who are married by education level and ethnic affiliation. The graphs reflect the higher marriage share in rural areas for both men and women, with little variation depending on the trait. Among women, those with less education tend to marry more often than their highly

[23] The authors use the following linguistic distance measure between two languages L_1 and L_2: $1 - ((d(L_1, R) + d(L_2, R) - d(L_1, L_2))/(d(L_1, R) + d(L_2, R))))^{\delta}$, where $d(L_i, R)$ is the length of path from a language to the tree's origin and $d(L_1, L_2)$ is the length of the shortest path from the first to the second language. The discount factor δ for short distances on the tree is set to its default value 0.5.

[24] Table 2.A.1 in Appendix 2.A presents a summary of these groupings by country.

Table 2.5.1: Summary of matching data

	Men		Women	
	Rural	Urban	Rural	Urban
Married	0.929	0.797	0.958	0.828
High education	0.605	0.853	0.452	0.758
Majority group	0.627	0.653	0.629	0.661
Observations	29705	15008	28791	14441

educated counterparts. This contrast becomes especially pronounced in the urban population, where the gap reaches 14 percentage points. For urban men, the difference between lower- and higher-educated groups is smaller but still notable at 6 percentage points. In comparison, disparities in marriage rates across ethnic groups remain low for both men and women in both rural and urban areas.

Figure 2.5.2 shows the shares of couples by spouses' traits within the married population.[25] In terms of education, homogamous couples are more prevalent than those with differing levels of education. This is consistent with patterns of educational assortative mating documented in the literature (Eika et al., 2019; Pesando, 2021). Yet, we see striking differences in rural and urban areas: in the urban population, the share of couples where both partners have higher education exceeds that of couples where both have lower education by 57 percentage points, while in rural areas the difference is only 4 percentage points. This may reflect both variations in preferences and in the distribution of education itself, as educational differences between men and women are more pronounced in urban settings, potentially causing a greater share of couples with different education levels.

With respect to ethnicity, similar assortative mating patterns emerge in both rural and urban settings. There is a larger share of couples with both spouses coming from a majority ethnic group. In urban areas, couples from non-majority groups are less common, while unions where only one spouse is from the majority group show a slightly higher level compared to rural areas.

2.5.3 Estimation

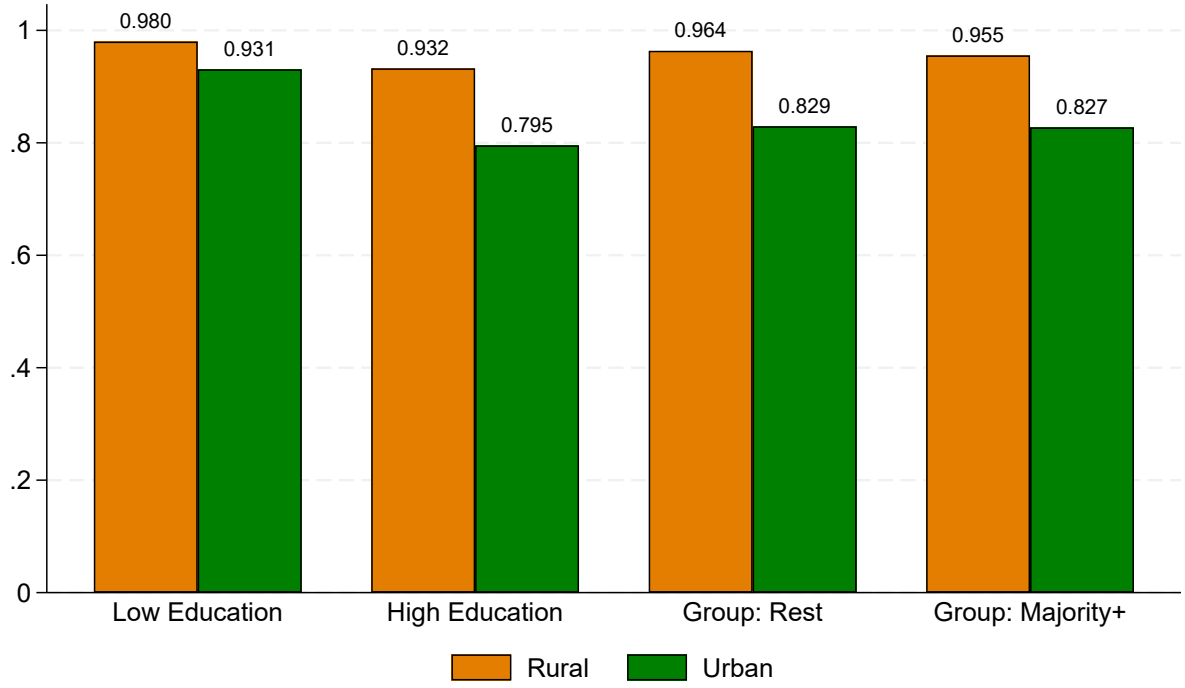
In order to gauge the relevance of matching on each dimension (education and ethnicity), we parametrize the marital surplus $\Phi(x, y)$ using a linear combination of basis functions:

$$\begin{aligned}\Phi(\beta) = \beta_1(Educ_x = Educ_y) + \beta_2(Ethn_x = Ethn_y) \\ + \beta_3(Educ_x = H) + \beta_4(Educ_y = H) \\ + \beta_5(Ethn_x = G_1) + \beta_6(Ethn_y = G_1)\end{aligned} \tag{2.5}$$

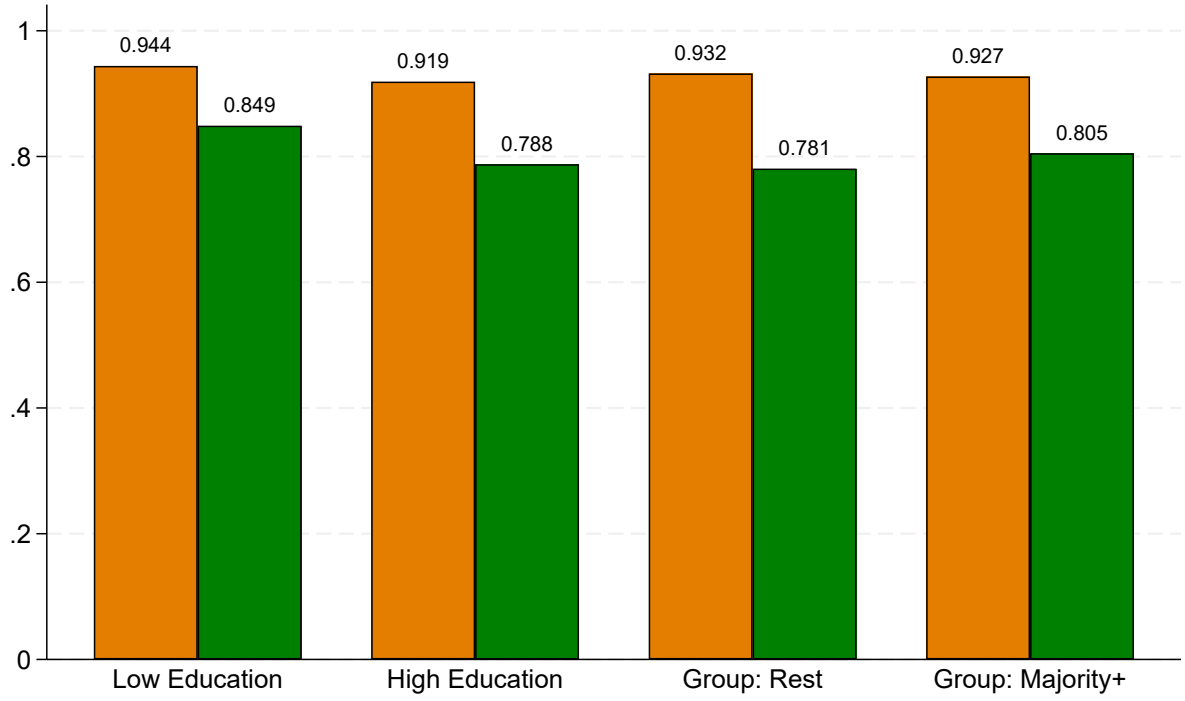
where parameters β_1 and β_2 indicate assortative mating preferences on education and ethnicity, respectively, and $\beta_3 - \beta_6$ contain the separate effects of education and ethnicity of the husband and the wife.

[25] Full contingency tables with numbers of men and women married or single by traits are presented in Appendix tables 2.A.2 and 2.A.3.

Figure 2.5.1: Marriage shares by education and ethnic group



(a) Women



(b) Men

The parameters are estimated using a Poisson GLM regression model with two-sided fixed effects following Galichon and Salanié (2024):

Figure 2.5.2: Shares of marriages by spouses' traits

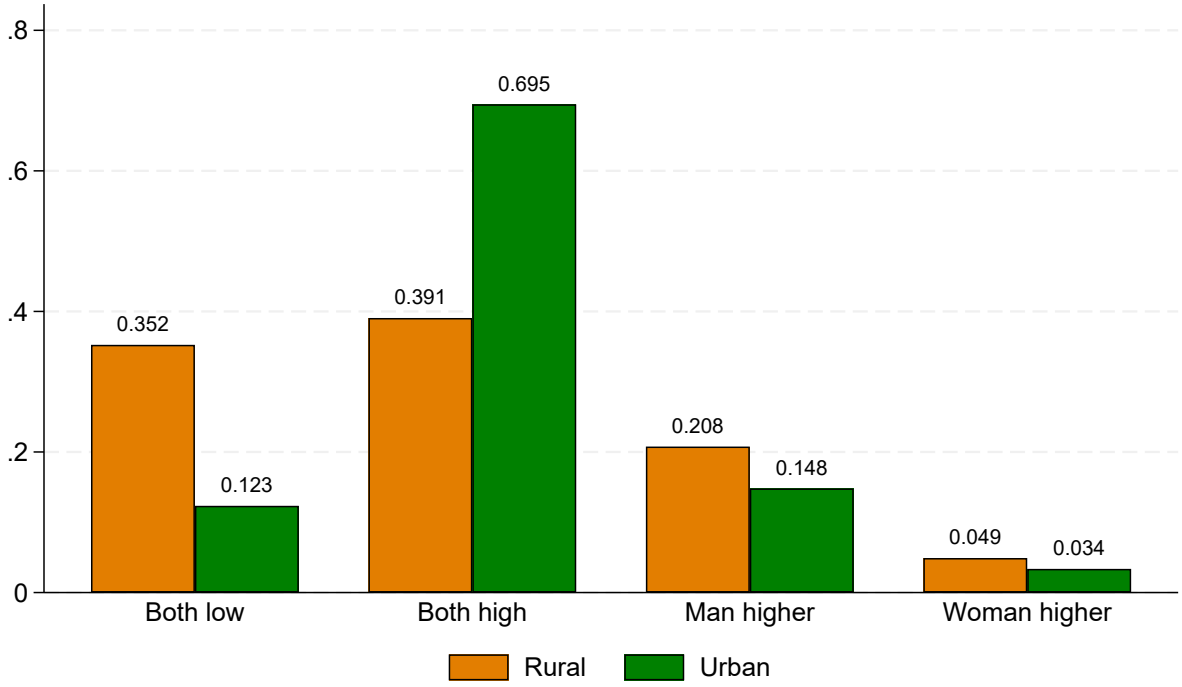


Figure 2.5.3: Education

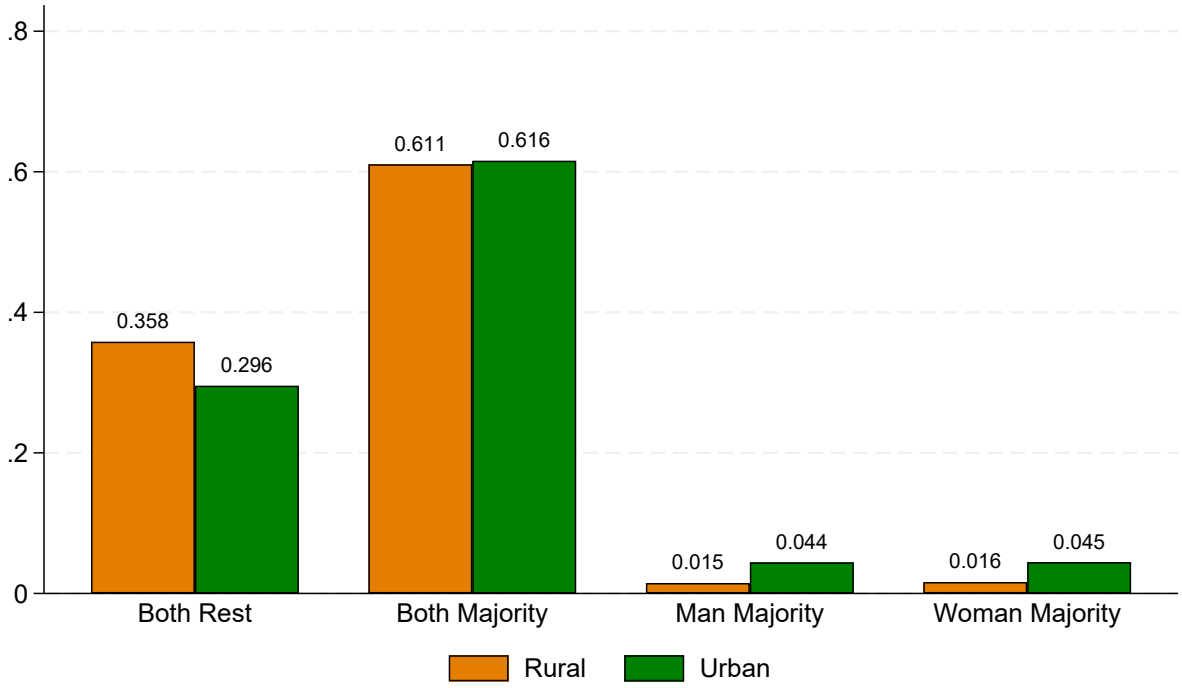


Figure 2.5.4: Ethnicity

$$\mathbb{E}(\mu|Z) = \exp\left(w\phi_{xy}\beta + w\mathbf{I}_X + w\mathbf{I}_Y\right) \tag{2.6}$$

where μ is a matrix of observed matches including singles, ϕ_{xy} is a matrix of basis functions with their respective parameters β as specified in equation 2.5, $\mathbf{I}_X$ and $\mathbf{I}_Y$ are indicator variables for types x and y, and w are weights that are 0.5 for couples and 1 for singles. As proven by Galichon and Salanié (2024), the estimate β in this Poisson regression is numerically equivalent to the estimate of a moment-matching estimator β using equation 2.5. Note that we estimate this model for the urban and rural market separately, i.e., the parameters are market-specific.

Results

Table 2.5.2 presents the parameters estimated using equation 2.5 for the rural and urban market. The full specification is presented in model (3), while models (1) and (2) show the effects of education and ethnicity separately. This allows us to compare how the two attributes interact in the market. Panel A contains the parameters of interest, β_1 and β_2, while Panel B shows the individual effects $\beta_3 - \beta_6$.

Table 2.5.2: Matching model estimation results

	(1)		(2)		(3)	
	Rural	Urban	Rural	Urban	Rural	Urban
A. Matching in attributes						
$Educ_x = Educ_y$	2.917***	2.738***			2.426***	2.542***
	(0.664)	(0.555)			(0.0714)	(0.171)
$Ethn_x = Ethn_y$			6.259***	3.677***	6.264***	4.141***
			(0.470)	(0.468)	(0.248)	(0.224)
B. Individual Attributes						
$Educ_x = H$	1.826**	0.778			0.234	-0.275
	(0.727)	(0.622)			(0.561)	(1.831)
$Educ_y = H$	-1.236	-2.278***			-3.728**	-4.325*
	(0.843)	(0.625)			(1.634)	(2.207)
$Ethn_x = G_1$			-0.490	-0.465	-0.357	0.104
			(0.694)	(0.988)	(0.482)	(0.319)
$Ethn_y = G_1$			-1.501*	-1.122	-0.620	0.0445
			(0.881)	(1.149)	(1.424)	(0.491)

Notes: Standard errors in parenthesis. $^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

The positive and significant results in Panel A show that matching on both attributes, education and ethnicity, is associated with utility gains in marriage in both rural and urban markets. This is true both when observing the two variables separately (models (1) and (2)) and jointly (model (3)). While matching on education increases marital gain by a similar amount in both rural (2.92) and urban (2.74) markets, matching on ethnicity increases utility in rural markets (6.26) more than in urban markets (3.68). This difference between the rural and the urban sample decreases slightly when considering both variables in model (3), due to the coefficient for ethnic matching increasing slightly in the urban sample from 3.68 in model (2) to 4.14 in model (3). This could imply that while in rural markets, preference for ethnic matching is independent of preferences for education, urban couples substitute more between education and ethnicity. Nevertheless, the ratio between the two coefficients is larger in the rural

sample (2.58) than in the urban sample (1.63). Interpreted as the marginal rate of substitution between the two attributes, this means that couples in rural areas would be more willing to trade off matching on education for matching on ethnicity, i.e., value ethnic matching more with respect to educational matching compared to urban couples.

With respect to the role of partners' individual attributes, Panel B shows that higher education of the husband tends to increase marital surplus in both rural and urban areas, though the standard errors on these coefficients are large. On the contrary, higher education of the woman decreases marital surplus in both samples. In other words, higher education of women presents itself as an unfavorable trait in marriage. The independent effects of husbands' and wives' ethnic group are less conclusive. The results in the rural sample suggest a small negative effect of any spouse belonging to the majority group. This would imply that couples where one or both spouses are from a majority group gain less utility from marriage compared to those where both spouses are of a non-majority ethnic group. For the urban sample, the coefficients turn from negative to positive when including education as an additional matching variable. However, the real value is hard to interpret as these effects tend to be small and cannot be statistically distinguished from zero. Rather, what seems to matter is the fact that both spouses belong to the same group.[26]

All in all, the estimation results show homogamy preferences both in education and ethnicity, in rural as well as urban marriages. Additionally, they suggest greater importance of ethnic matching for rural couples. To further investigate how these preferences affect the incidence of interethnic marriages in urban areas, we refer to a counterfactual analysis in the following section.

2.5.4 Counterfactual analysis

The utility estimation results show that preferences for endogamous marriages are higher in rural areas than in urban areas. We now explore how three channels affect interethnic marriages in urban areas: (1) the distribution of ethnic groups, (2) the overall structure of the marriage surplus, and (3) assortative mating preferences in ethnicity. To do this, we employ a counterfactual analysis where we set either the surplus or the population supplies in the urban market equal to the rural market as described below. With this exercise, we decompose how much of the difference in intermarriage between rural and urban areas is due to the different channels.

For all three exercises, we compute the counterfactual matching patterns by combining the model result (2.4) with the market constraints (2.2) to yield a system of equations:

$$N(x) = \sum_y \sqrt{\mu(x,0)\mu(0,y)} \exp\left(\frac{\Phi(x,y)}{2} + \mu(x,0)\right) \tag{2.7}$$

$$M(y) = \sum_x \sqrt{\mu(x,0)\mu(0,y)} \exp\left(\frac{\Phi(x,y)}{2} + \mu(0,y)\right) \tag{2.8}$$

that can be solved iteratively for the equilibrium masses of singles, $\mu(x,0), \mu(0,y)$. The

[26] Table 2.A.4 in Appendix 2.A repeats the estimation without migrants, i.e., with a sample of men and women that have always lived in their current residence. The matching coefficients remain stable and have similar magnitude, thus providing similar conclusions.

algorithm developed by Galichon and Salanié (2022) used to solve these simultaneous equations is presented in Appendix 2.C.

Ethnicity distributions In the first counterfactual, we want to keep the preference structure intact in the urban market, but adjust the population sizes in order to mimic the distribution of ethnic groups in the rural market. We achieve this by predicting matches given the observed joint utilities $\Phi^U(x, y)$ in the urban market U and a counterfactual distribution of ethnic groups in the urban market. In particular, we adjust the numbers of men and women in the urban market by ethnic group (summed over education level) such that they correspond to the rate at which they are observed in the rural market, i.e.:

$$\tilde{N}^U(Ethn_x) = Share^R(Ethn_x) \times N^U$$
$$\tilde{M}^U(Ethn_y) = Share^R(Ethn_y) \times M^U$$

where $N^U = \sum_x N^U(x)$ and $M^U = \sum_y M^U(y)$ are the total number of observed men and women in the urban market, respectively. We create a counterfactual matrix of men and women across education levels and ethnic groups such that the constraints of (observed) number by education level and (counterfactual) number by ethnic groups are satisfied. For example, for men we compute:

Table 2.5.3: Counterfactual matrix of men's types

	Majority+	Rest	Total Row
Low education	$\tilde{N}^U(x = LG_1)$	$\tilde{N}^U(x = LG_2)$	$N^U(Educ_x = L)$
High education	$\tilde{N}^U(x = HG_1)$	$\tilde{N}^U(x = HG_2)$	$N^U(Educ_x = H)$
Total Column	$\tilde{N}^U(Ethn_x = G_1)$	$\tilde{N}^U(Ethn_x = G_2)$	N^U

given $N^U(Educ_x)$ and $\tilde{N}(Ethn_x)$ and equivalently for women:

Table 2.5.4: Counterfactual matrix of women's types

	Majority+	Rest	Total Row
Low education	$\tilde{M}^U(y = LG_1)$	$\tilde{M}^U(y = LG_2)$	$M^U(Educ_y = L)$
High education	$\tilde{M}^U(y = HG_1)$	$\tilde{M}(y = HG_2)$	$M^U(Educ_y = H)$
Total Column	$\tilde{M}^U(Ethn_y = G_1)$	$\tilde{M}^U(Ethn_y = G_2)$	M^U

The counterfactual numbers of men and women by education-ethnicity type are found using an iterative proportional fitting procedure (Deming & Stephan, 1940).

Overall marital surplus The second counterfactual aims at changing the preference structure in the urban market to that of the rural market while holding constant the marginal distributions of men and women observed in the urban market. We obtain the counterfactual matching by setting the entire marital surplus in the urban to that of the rural market, $\tilde{\Phi}^U(x, y) = \Phi^R(x, y)$ for any type of couple. This encapsulates both preferences within marriages as well as the overall gains from marriage, or staying single as the outside option.

Endogamy preferences The last exercise consists of quantifying the effect of different preferences for endogamous marriages between rural and urban marriage markets. For this, we first decompose the marital surplus in market $m = \{R, U\}$ into the degree of assortative mating in education and ethnicity and the overall gains from marriage for a man type x and woman type y. We then fix the degree of assortative mating $\phi^U(Ethn_x = Ethn_y)$ in the urban market to that of the rural market $\phi^R(Ethn_x = Ethn_y)$ to enter into the urban surplus Φ^U_{xy}:

$$\tilde{\Phi}^u_{xy}(Ethn, Educ) = \phi^r(Ethn_x = Ethn_y) + \phi^u(Educ_x = Educ_y) + a^u_x + b^u_y$$

The values a_x and b_y are chosen such that the counterfactual marriage rates match the observed marriage rates, i.e., the number of singles in the market remain constant. Following Dupuy and Weber (2022), we effectively use a matching model without singles to find equilibrium values of a_x and b_y to achieve this. The algorithm outlined in Appendix 2.C explains this procedure.

Simulation Results

Figure 2.5.5 presents the main results of the counterfactual exercises described above. The graph displays the share of intermarried couples in rural and urban areas in the baseline scenario, i.e., under observed preferences, compared to the counterfactual predictions in the urban market. As the model slightly over-predicts the number and share of interethnic marriages in both rural and urban areas[27], we present both the observed and predicted intermarriage shares to allow better comparability with the counterfactuals.

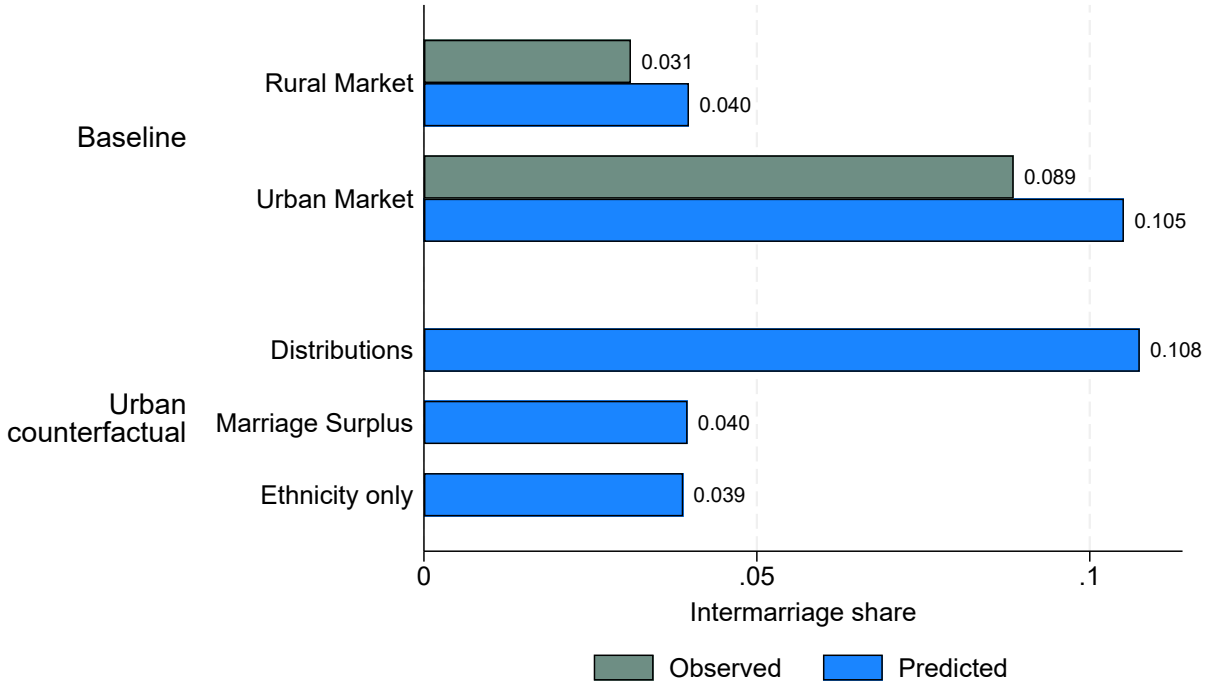


Figure 2.5.5: Intermarriage shares under baseline and counterfactual scenarios

Comparing the actual urban intermarriage rate with that under the counterfactual

[27] See Figure 2.B.1 in Appendix 2.B for details on the model fit.

with group shares equal to the rural market ("distributions"), the level is factually unchanged. The difference in the observed and counterfactual shares comes purely from unexplained heterogeneity in the model, as the baseline and counterfactual predictions are virtually the same.

By contrast, the intermarriage rate drops substantially from 10.5% (predicted) to 4% under the scenario where total marriage surplus is equal to that of the rural market. This number corresponds to the predicted share of intermarried couples in rural areas and differs only by 1 percentage point from the observed rural intermarriage rate. This indicates that marital preferences explain a large share of the difference between intermarriage rates in urban and rural markets. In other words, if the urban population would have the same incentives to marry as the rural population, they would select into same-ethnicity marriages at the same rate as their rural counterparts.

A similarly large change is recorded in the last counterfactual, where marriage rates are held constant but the utility from matching on ethnicity is adjusted to the rural level ("ethnicity only"). Again, the intermarriage share in the counterfactual urban market is very similar to both the predicted and observed share in rural areas. This implies that ethnicity preferences explain the most part of the effect of changes in marital surplus, whereas selection into marriage do not play a role in changing the propensity to intermarry.

To see which types of men and women drive these results, we further disentangle the selection into marriages under the counterfactual marriage surplus (the second counterfactual) in Table 2.5.5. The table illustrates how men and women sort into marriage and, conditional on marrying, into specific types of marriage – marrying someone with the same education level or ethnic group – according to their own education and ethnic affiliation. While the marriage rate increases for all types, lower educated men and women only marginally adjust their choice to marry. On the contrary, those with higher education are now relatively more inclined to marry. This reflects higher incentives to marry under rural preferences, especially for those with a higher education. No large differences emerge in terms of the ethnic background, as both groups are around 10 percentage points more likely to select into marriage under the counterfactual utility. Notably, the selection into endogamous marriages rises by magnitudes more than educationally homogamous marriages. For lower-educated men and women, as well as a small share of higher educated women, this even leads to a decrease in couples with the same education level. This result is consistent with a trade-off between matching on education and on ethnicity, where ethnic matching is favored at the expense of finding a spouse with the same education level.

These results suggest that preferences, for marriage and especially for ethnicity, are more important in explaining differences in interethnic marriages between rural and urban areas than are the population supplies of different ethnic groups across these marriage markets. The first result, that urban markets hold less overall marital surplus that lead to less marriages, especially for higher educated men and women, is in line with findings in Luke and Munshi (2006). Their study in Kenya suggests that having better labor market options in cities reduces the importance of marriage to tap into family networks. On top of this, we find that differences in ethnic assortative mating preferences play an important role in explaining rural-urban differences in interethnic marriage, and that preferences for education enter as a competing preference in urban settings.

Table 2.5.5: Changes in urban marriage selection by type and gender (with rural marriage surplus)

	Men			Women		
Attribute	Marriage	Same Education	Same Ethnicity	Marriage	Same Education	Same Ethnicity
Education						
Low	1.59	-4.35	7.00	0.37	-2.08	7.06
High	11.25	1.65	6.45	13.35	-0.15	6.34
Ethnic group						
Majority+	9.59	1.10	5.28	10.27	0.79	4.84
Rest	10.28	0.28	8.90	10.10	0.80	9.71

Notes: The table shows percentage point changes in marriage selection between predicted baseline and counterfactual matching, where the marriage surplus correspond to those in the rural market. A positive value indicates that, for the given attribute, a larger share of people with the corresponding attribute select into marriage under the counterfactual than under the observed matching. The column "marriage" collects the total change in marriage selection, and the columns "same education" and "same ethnicity" indicate matching with someone of the same education level or ethnic group.

2.6 Conclusion

This study investigates the link between urbanization and interethnic marriage in Sub-Saharan Africa. Using harmonized cross-country data, we find a positive correlation between the urbanization rate and interethnic marriage both across countries and over marriage cohorts. Exploiting localized variation in urbanization over time, we further show that this positive relationship is robust and increases with the level of urbanization. Our structural approach to investigate the channels of this effect further sheds light on the role of marital preferences and population sizes of ethnic groups in urban and rural areas. Using a matching model and observed marriages based on education and ethnicity, we document assortative mating on education and ethnicity in both rural and urban marriage markets. However, the results suggest that matching on ethnic identity plays a larger role in rural markets compared to cities. To further identify the effect of marital preferences versus market-level group sizes, we simulate marriage equilibria based on counterfactual scenarios. These findings highlight the importance of the utility in marriage, especially driven by ethnicity preferences, to explain differences in interethnic marriage rates between urban and rural areas.

Our exploration provides novel quantitative evidence of how cultural and socioeconomic preferences interact in the marriage market. It is, however, limited in providing explanations on *how* preferences are being changed in urban settings. Our results suggest that the relative importance of education plays a role in shifting preferences. This could be due to increased intergroup contact in schools or workplaces as well as the increasing relevance of socioeconomic status. Preferences are also likely to interact with social norms and traditions. For example, rural unions could be more likely to be guided by family involvement and thus represent the preferences of the family or even the larger community. Future research may address which of these explanations prevail and the specific conditions under which they apply.

With ethnic diversity in Sub-Saharan Africa often being associated with heightened conflict, this study shows that urban environments can foster greater openness to non-

traditional marriages, and can thereby promote mututal understanding and acceptance.

Appendix

2.A Tables

Table 2.A.1: Summary of aggregated groups for matching

Country	Majority Group	M groups	R groups	Share M groups
Benin	fon/related	2	6	0.597
Burkina Faso	mossi	3	6	0.761
Central African Republic	gbaya	6	3	0.813
Côte d'Ivoire	akan	1	1	0.716
Cameroon	bamilike/central	15	29	0.404
Congo	balari	54	3	0.988
Ethiopia	oromo	4	4	0.442
Ghana	mole/dagbani	3	8	0.463
Guinea	malinke	2	4	0.435
Kenya	kalenjin	1	10	0.515
Mali	bambara	2	7	0.349
Mozambique	cisena/simili	5	1	0.961
Niger	haoussa	1	7	0.477
Senegal	wolof/lebou	3	4	0.855
Sierra Leone	mende	2	6	0.478
Chad	sara	3	4	0.433
Togo	kabye/tem	2	3	0.642
Uganda	baganda	19	14	0.705
Zimbabwe	black	1	2	0.982

Notes: Column "M groups" lists the number of groups categorized within "M" (majority or similar), column "R groups" lists the number of groups categorized within "R" (rest).

Table 2.A.2: Education contingency table

	Rural			Urban		
Women:	Low	High	$\emptyset$	Low	High	$\emptyset$
Men:						
Low	17712	2096	5613	2546	608	1538
High	8619	14171	20145	2585	10016	18158
$\emptyset$	2666	7794		1234	9821	

Notes: Table presents the number of men and women married or single by education level. Women's levels are presented in the columns and men's in the rows. "Low" refers to no formal education, and "high" refers to primary level or higher. The symbol $\emptyset$ represents the singles from the opposite sex.

Table 2.A.3: Ethnicity contingency table

	Rural			Urban		
Women:	Rest	Majority+	$\emptyset$	Rest	Majority+	$\emptyset$
Men:						
Rest	15460	725	9726	4820	652	6915
Majority+	667	25746	16032	689	9594	12781
$\emptyset$	3740	6720		3691	7364	

Notes: Table presents the number of men and women married or single by ethnic group. Women's levels are presented in the columns and men's in the rows. "Majority+" refers to affiliation with majority or similar groups, and "Rest" refers to all other groups. The symbol $\emptyset$ represents the singles from the opposite sex.

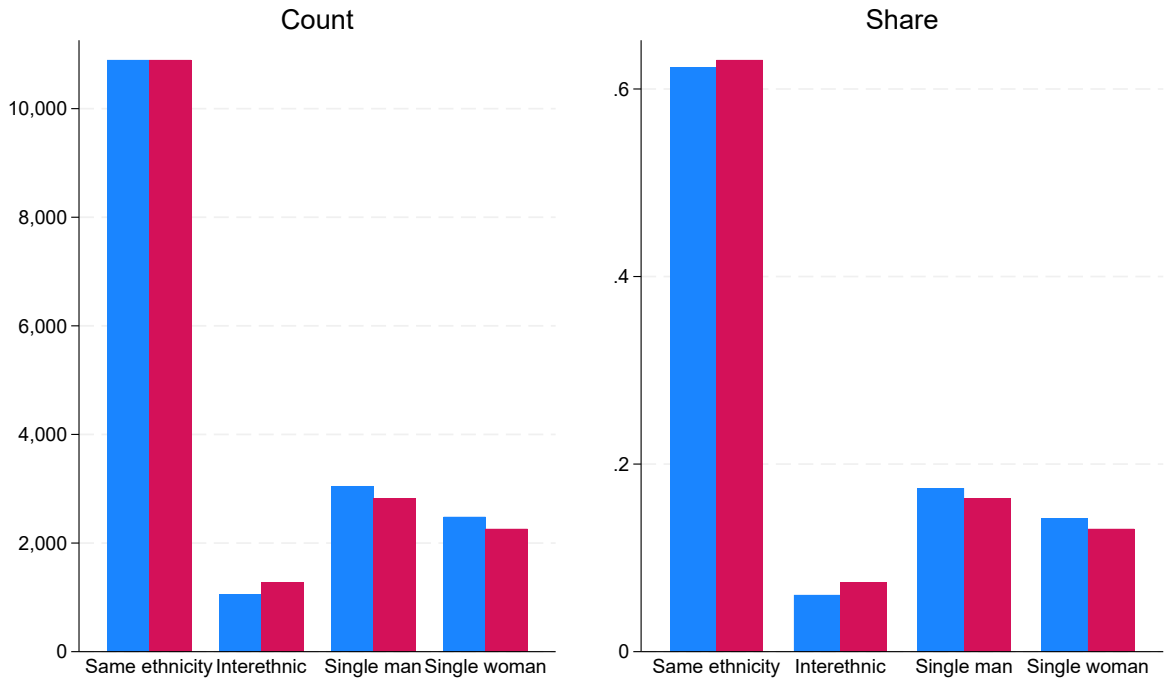
Table 2.A.4: Robustness of matching estimation to exclusion of migrants

	(1)		(2)		(3)	
	Rural	Urban	Rural	Urban	Rural	Urban
A. Matching in attributes						
$Educ_x = Educ_y$	2.892***	2.284***			2.200***	2.061***
	(0.658)	(0.562)			(0.159)	(0.291)
$Ethn_x = Ethn_y$			5.889***	2.517***	5.953***	3.597***
			(0.683)	(0.784)	(0.466)	(0.409)
B. Individual Attributes						
$Educ_x = H$	0.971	-0.831			-0.446	-1.630
	(0.719)	(0.752)			(0.735)	(1.532)
$Educ_y = H$	-2.580***	-3.364***			-4.235**	-3.995***
	(0.822)	(0.767)			(1.805)	(1.547)
$Ethn_x = G_1$			-0.953	-2.079	-0.516	-0.607
			(0.949)	(1.506)	(0.703)	(0.595)
$Ethn_y = G_1$			-2.188	-1.820	-1.102	-0.946
			(1.349)	(1.394)	(1.683)	(0.630)

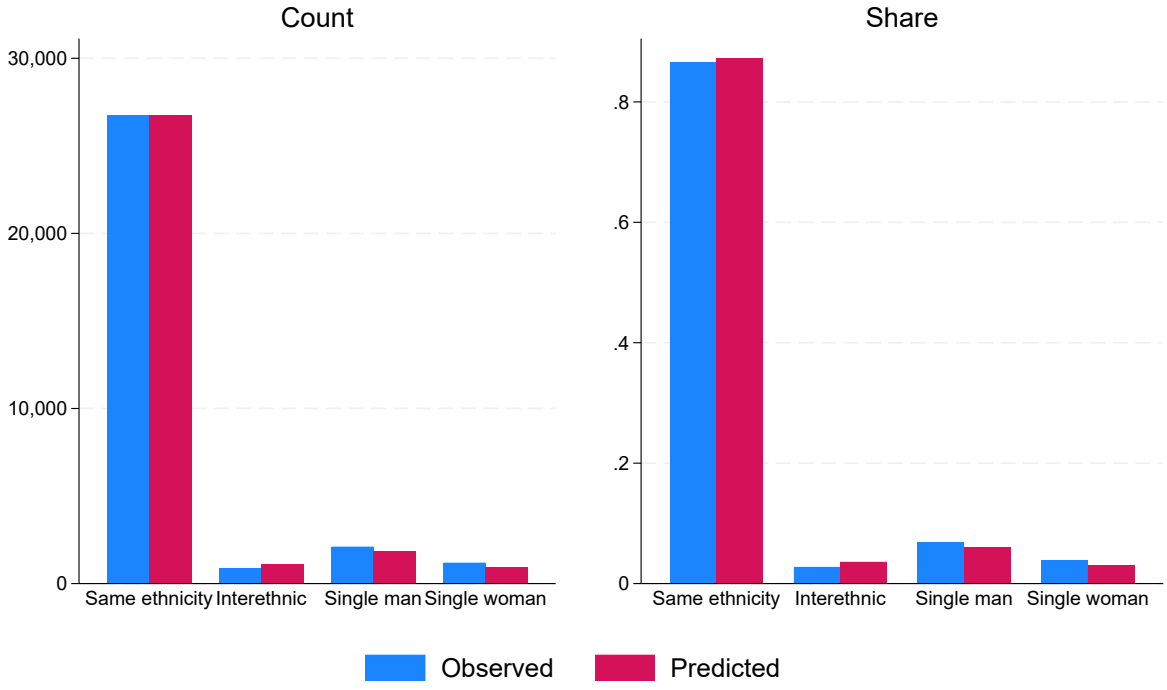
Notes: Standard errors in parenthesis. $^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$. For this robustness check, we only include men and women that indicate that they have always lived in their current residence.

2.B Figures

Figure 2.B.1: Model Fit: Observed vs. predicted matching



(a) Urban



(b) Rural

Figure 2.B.1 shows the goodness of fit of model 2.6 by comparing the observed

and predicted couples and singles in the urban and rural market. While there is a close fit between the observed and predicted matches, the model systematically over-predicts interethnic marriages in both rural and urban markets at the expense of under-predicting single men and women. The endogamous marriages (same ethnicity) are perfectly matched in counts due to the model parameters, whereas they are slightly over-predicted in shares.

2.C Algorithms

Algorithm with singles:

Rewrite equations 2.7 and 2.8 to:

$$\mu(x,0) = \left(\sqrt{N(x) + K(x)^2} - K(x)^2\right) \tag{2.9}$$

$$\mu(0,y) = \left(\sqrt{M(y) + P(y)^2} - P(y)^2\right) \tag{2.10}$$

where $K(x) = \exp\left(\frac{\Phi(x,y)}{2}\sum_y \sqrt{\frac{\mu(0,y)}{2}}\right)$ and $P(y) = \exp\left(\frac{\Phi(x,y)}{2}\sum_x \sqrt{\frac{\mu(x,0)}{2}}\right)$. Start with an initial value for $\mu(0,y)$ and solve for expression 2.9. Plug in the new value of $\mu(x,0)$ into expression 2.10 and solve for the new $\mu(0,y)$. Continue the loop until convergence to the new equilibrium matching $\{\mu^*(x,0), \mu^*(0,y), \mu^*(x,y)\}$ where

$$\mu^*(x,y) = \exp\left(\frac{\Phi(x,y)}{2}\sqrt{\mu^*(x,0)\mu^*(0,y)}\right)$$

Algorithm without singles:

Using the margins $\dot{N}(x) = N(x) - \mu(x,0)$ and $\dot{M}(y) = M(y) - \mu(0,y)$, as well as the joint utility $\tilde{\Phi}^u(x,y)$, we rewite equations 2.7 and 2.8 as:

$$a_x = 2\log\left(\frac{\dot{N}(x)}{\sum_y \exp\left(\frac{\tilde{\Phi}^u(x,y)+b_y}{2}\right)}\right) \tag{2.11}$$

$$b_y = 2\log\left(\frac{\dot{M}(y)}{\sum_x \exp\left(\frac{\tilde{\Phi}^u(x,y)+a_x}{2}\right)}\right) \tag{2.12}$$

Similar to above, we choose an initial value of b_y and plug it into equation 2.11 solving for a new a_x. Using the new value of a_x, we solve for b_y and continue this iterative procedure until convergence to equilibrium values $\{a_x^*, b_y^*\}$. The equilibrium matching (without singles) is solved as $\mu^*(x,y) = \exp\left(\frac{\tilde{\Phi}^u(x,y)+a_x^*+b_y^*}{2}\right)$.

Chapter 3

Attitudes towards Immigration in a Highly Multicultural Society: The Roles of Foreign Background and Local Exposure

3.1 Introduction

Long-term international comparisons reveal significant changes in the size and structure of immigration to developed countries. Since 1960, the foreign-born population grew much faster than the total population, significantly increasing the average share of immigrants in rich countries (Migration Policy Institute, 2024). The causes of this trend are well documented and include large disparities in population growth between rich and poor countries, significant economic inequalities, increased ease of travel due to globalization, political instability, climate shocks, and other factors. Immigration from geographically, economically, politically, and culturally distant countries is a widespread phenomenon, especially in metropolitan areas, which can influence attitudes toward immigration (Bauer et al., 2000; Facchini & Mayda, 2017; Mayda, 2006) and political preferences (Docquier et al., 2024; Guriev & Papaioannou, 2022).

Migration trends in the Grand Duchy of Luxembourg are even more striking than in other industrialized countries. Between 1960 and 2020, the share of first-generation immigrants increased from 13.1% to 47.1%, and the share of immigrants from developing countries increased from 1.3% to 10.4%. Recent trends have been driven by significant inflows from several countries, including China, Montenegro, India, Cape Verde, Kosovo, Brazil, Russia, Morocco, and Turkey (Docquier et al., 2023). Hence, cultural diversity has been rising in recent decades, and first-generation immigrants, especially recent arrivals, are disproportionately concentrated at both ends of the skill and income distribution (Verheyden et al., 2024). In addition, second-generation immigrants make up about a quarter of the total population (Docquier et al., 2023). In this highly multicultural context, attitudes towards immigration can have a significant impact on the quality of life, social cohesion and the ability to live together harmoniously within a society.

We empirically investigate the individual factors that shape attitudes toward immigration in Luxembourg, as well as the conditions under which positive or negative views emerge. Our analysis is based primarily on the *National Survey on Racism and Ethno-Racial Discrimination in Luxembourg*, which we conducted online in the summer of 2021 (Bienvenue et al., 2021). The survey targeted individuals aged 18 and older who were living in the Grand Duchy of Luxembourg in June 2021, including both native and immigrant respondents, disaggregated by country/region of origin. Participants were asked about their views on the impact of immigration on the economy, culture, and quality of life in Luxembourg. Further, respondents could voice their opinions about the perceived level of immigration and their expectations regarding language integration. These perceptions are the dependent variables in our empirical analysis. The survey also provides extensive information on respondents' socio-demographic characteristics, immigrant backgrounds, and stereotypes about some groups at risk of discrimination, such as Muslims, Blacks, or Eastern Europeans.

We were also able to use national Social Security records to accurately determine the composition of respondent's neighborhood in terms of birthplaces,[28], allowing us to proxy their exposure to different types of immigrants, which Card et al. (2012) describe as compositional amenities. We take a flexible approach to defining the neighborhood, considering various definitions ranging from the population living on a single street to those living on nearby streets, with centroids located within 0.5, 1, or 3 kilometers from

[28] As described in Section 3.2 around 66% of the initial survey respondents consented to linking their information from the Social Security records.

the residential street. The definition of the neighborhood is crucial in determining the exogeneity of its composition, especially given the intense competition for housing in the Luxembourg housing market. We consider the composition of the neighborhood to be quasi-random, since it is extremely difficult for individuals to choose their exact place of residence. Extending the neighborhood definition to a larger area may provide a more accurate measure of local exposure to immigration, but at the cost of increasing endogeneity concerns. By comparing neighborhoods of different sizes and employing an instrumental variable, we explore the causal impact of immigration and personal contact with immigrants on attitudes (in line with Allport, 1954; Paluck et al., 2019; Pettigrew & Tropp, 2006).

Our study provides valuable insights into the complex interplay of factors that shape attitudes towards immigration in a highly diverse and multicultural society such as Luxembourg. Unlike many other studies, we go beyond the native population and include foreign-born respondents and second-generation immigrants in our analysis, providing a more comprehensive understanding of attitudes toward immigration (see also Gihleb et al., 2022). Our data also allow us to assess local exposure to immigration at different spatial scales, exploring the role of migrants' origin, and to interact neighborhood composition variables with individual sociodemographic and economic characteristics of respondents. Overall, our study offers a nuanced view of the factors influencing attitudes towards immigration, which is crucial for promoting social cohesion and living together in a multicultural society.

We show that a large majority of respondents acknowledge that immigration enriches Luxembourg's national identity and has a positive impact on the economy. A majority of natives have positive views of immigration, a sentiment that may be influenced by trust in institutions (Sarracino & Riillo, 2022) as well as observed labor market complementarities, particularly between the private and public sectors (Verheyden et al., 2024). The latter is consistent with Bauer et al. (2000), who compare attitudes toward immigration in 12 OECD countries and find that natives view immigration more favorably in countries where immigrants are selected according to labor market needs. In line with other studies (Drazanova et al., 2022; Facchini & Mayda, 2017; Hainmueller & Hopkins, 2014), we also find that these attitudes are correlated with sociodemographic characteristics such as age and education. More interestingly, the foreign background of the respondent is a crucial determinant of attitudes. Portuguese immigrants – who make up the largest share of immigrants in Luxembourg – have the most positive attitudes towards immigration, followed by non-EU immigrants and immigrants from neighboring countries. Second-generation immigrants behave similarly to the latter and are more supportive of immigration than other natives.

In addition, the total share of immigrants in the immediate neighborhood, regardless of origin composition, does not significantly affect views whether immigration benefits the economy or enriches national identity. However, local exposure to recent immigration negatively affects natives' perceptions of the current level of immigration. This may suggest that acceptance of diversity and multiculturalism takes time (as in Levi et al., 2020). These findings contrast with many studies conducted on refugees or ethnic minorities in other countries (Schneider-Strawczynski, 2021; Steinmayr, 2021), confirming that attitudes towards immigrants are likely to be determined by the perceived contribution of immigrants to human capital accumulation and economic performance.

Our study contributes to the existing literature on the political economy of immigration, which examines the relationship between immigration and attitudes towards

immigrants (e.g. Bauer et al., 2000; Hainmueller & Hopkins, 2014; Mayda, 2006), attitudes towards redistribution (e.g. Alesina et al., 2021, 2023; Moriconi et al., 2019), and the effect of immigration on voting for populist parties in different countries (e.g. Docquier et al., 2024; Guriev & Papaioannou, 2022). Both economic and political science studies have consistently shown that attitudes toward immigration are influenced by individual-level characteristics such as education (positively) and age (negatively), and are motivated by concerns about the negative labor market or fiscal effects of immigration (Facchini & Mayda, 2017) or identity/cultural issues (Card et al., 2012; Hainmueller & Hopkins, 2014). Consequently, attitudes vary quantitatively across contexts (Drazanova et al., 2022), depending on the average characteristics of the immigrant population, especially the skill structure (Edo et al., 2019; Hainmueller & Hiscox, 2010; Moriconi et al., 2019; Moriconi et al., 2022). In Luxembourg, over 80% of the university educated workforce were born abroad. Taking into account cross-border commuters, this proportion rises to 90% (Verheyden et al., 2024).

As noted above, our study differs from previous work in that our sample is not limited to the native-born, allowing us to examine the role of foreign background on attitudes toward immigration. Our approach is in line with recent studies such as Bursztyn et al. (2024) and Domènech-Arumí (2025), who examine how the presence of individuals of a particular foreign origin affects perceptions of immigration and preferences for redistribution, and Gihleb et al. (2022), who examine attitudes toward newcomers among natives and previous immigrants. By examining the role of foreign background in shaping attitudes toward immigration, our study makes a unique and valuable contribution to the political economy of immigration literature.

We pay particular attention to the birthplace composition of the immediate neighborhood. We draw on intergroup contact theory originally proposed by Allport (1954), which suggests that frequent interpersonal contact between majority and minority group members can reduce prejudice under certain conditions. This hypothesis has been tested in many contexts, with meta-analyses indicating that contact generally reduces prejudice, but the magnitude of the effect varies depending on the type of heterogeneity (e.g., ethnic versus racial) between people or the level of diversity in the population (Paluck et al., 2019; Pettigrew & Tropp, 2006).

Several studies have tested contact theory in different contexts. For example, Corno et al. (2019) used a random roommate assignment policy at a large South African university to show that living with a roommate of a different race reduces negative stereotypes of blacks among white students and increases interracial friendships. Meanwhile, Bazzi et al. (2019) used a population resettlement experiment in Indonesia to show that migrant integration increases in ethnically fractionalized communities with many small groups, but is smaller in polarized communities or areas with strong residential segregation. Recent studies have also tested the contact theory by examining the exogenous assignment of asylum seekers to municipalities. Consistent with contact theory, Steinmayr (2021) found that mere exposure to refugees passing through border municipalities increased far-right votes in Austria, while contact and sustained interactions between natives and asylum seekers in host municipalities decreased far-right votes. In contrast, Schaub et al. (2021) found that anti-immigrant sentiments were unaffected by the presence of refugees in respondents' municipalities in rural eastern Germany. Similarly, the effect of the local share of immigrants on attitudes and political preferences varies across contexts (Cools et al., 2021; Guriev & Papaioannou, 2022). Since it takes time to establish contacts, we also examine whether attitudes respond differently to the total stock and recent influx of immigrants in the neigh-

borhood, allowing us to explore the dynamic nature of attitudes towards immigration (Levi et al., 2020).

Our study contributes to this literature by examining different measures of local exposure, including immigrants from different origins as well as alternative measures of the neighborhood. By examining the role of compositional factors on attitudes toward immigration, our study provides valuable insights into the mechanisms underlying intergroup contact theory and sheds light on the conditions under which exposure to immigrants can reduce prejudice.

The following sections of our paper are organized as follows. Section 3.2 provides a comprehensive overview of our data sources and highlights some key stylized facts. Section 3.3 describes our empirical strategy. Section 3.4 presents the main results of our analysis, focusing on the determinants of attitudes toward immigration and the role of neighborhood exposure. Finally, section 3.5 concludes and discusses the policy implications of our findings.

3.2 Data sources and stylized facts

We combine data from the *National Survey on Racism and Ethno-Racial Discrimination in Luxembourg* (Bienvenue et al., 2021), which we conducted online in the summer of 2021, and the social security database maintained by the IGSS.[29] IGSS data provide administrative information on the population living or working in the country. The target population of the survey was people aged 18 and over. During the survey, respondents were asked to provide access to their social security records. In total, 1,954 respondents gave their consent out of 2,949 individuals who completed the questionnaire.[30] The survey responses of these respondents were then linked to their administrative data available in the IGSS to produce the database used in the project. This section briefly presents key stylized facts about attitudes toward immigration and neighborhood composition. Table 3.B.1 in the Appendix 3.B reports descriptive statistics for the primary variables used in our study.

The survey was designed to capture attitudes towards immigration using four statements, presented to individuals aged 18 years and older who were resident in Luxembourg. The four statements are as follows: "the presence of immigrants enriches the identity of the country" (S1); "Luxembourg needs immigrants to sustain its economy and social protection system" (S2); "there are too many immigrants in Luxembourg" (S3); and "all immigrants should learn to speak Luxembourgish" (S4). These statements were chosen to capture different dimensions of attitudes towards immigration. Specifically, S1 is intended to elicit perceptions of cultural benefits from immigration, while S2 captures perceptions of economic benefits. S3 reflects respondents' views on the optimal level of immigration to the country, while S4 expresses the population's expectations of immigrants' linguistic assimilation.

Figure 3.2.1 shows the distribution of respondents' opinions for each statement: including full agreement, partial agreement, partial disagreement, full disagreement, or no response – due to either a preference not to answer or a lack of knowledge. Our analysis focuses on identifying factors that influence *pro-immigration attitudes*, as shown by the green bars. On average, 83.5% of respondents think that immigration

[29] The General Inspectorate of Social Security, or in French, l'Inspection générale de la sécurité sociale.

[30] Appendix 3.A discusses the similarities and differences between the full and the consent sample in terms of sociodemographic background characteristics and attitudes.

strengthens national identity, 81.5% think that immigration has a positive impact on the economy, 54.5% do not think that the number of immigrants is too high, while 29.3% do not think that all immigrants should be forced to learn Luxembourgish. These percentages refer to the total pool of respondents, including those who did not give a definite answer. Overall, the Luxembourg population has a strong positive attitude towards the cultural and economic benefits associated with immigration. However, there are mixed views on the optimal level of immigration, and a majority of respondents believe that immigrants should be able to speak the local language.[31]

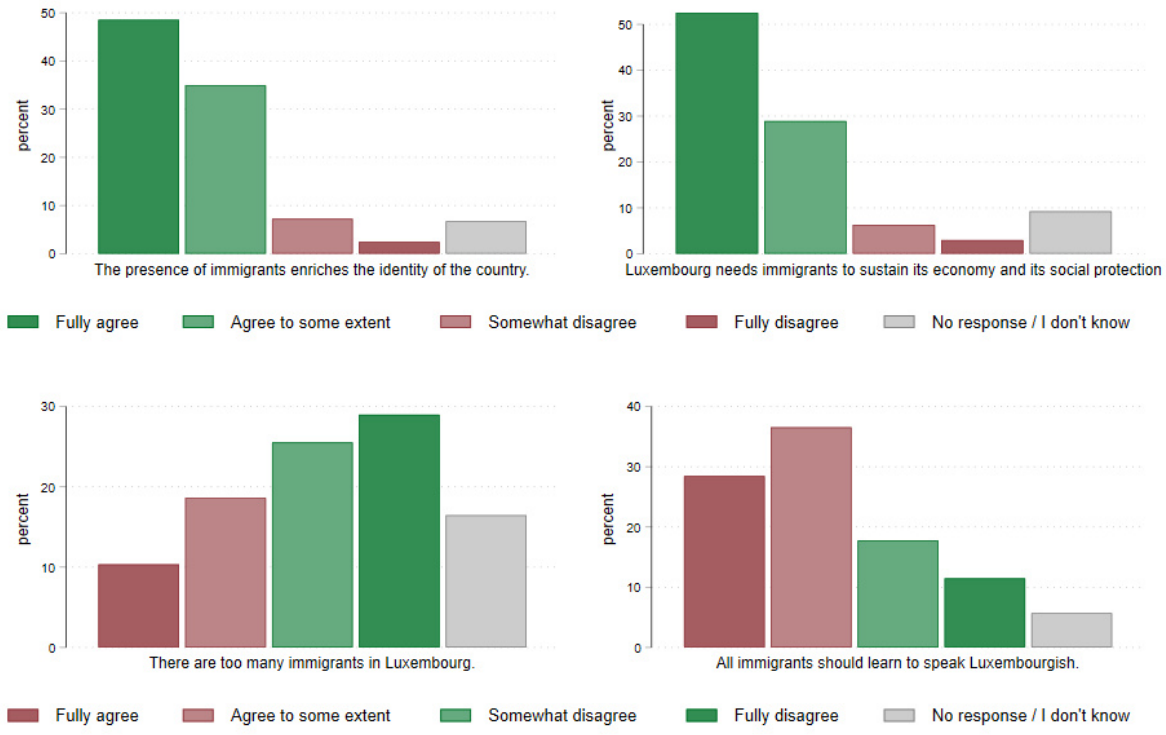


Figure 3.2.1: Attitudes toward immigration: Responses to the national survey

Given our focus on *pro-immigration attitudes* in the empirical analysis, we use dependent variables that take the value of one if the respondent agrees with the first two statements, and disagrees with the last two statements. We establish four binary variables that capture individuals' favorable stances towards immigration, denoted as follows:[32]

1. ENI (*immigration Enriches National Identity*) takes a value of one when the respondent fully or partially agrees with statement S1; otherwise, it is equal to zero.
2. MEB (*immigration Makes the Economy Better*) takes a value of one when the respondent fully or partially agrees with statement S2; otherwise, it is equal to zero.

[31] Luxembourgish is one of three official languages used in the country, including French and German.

[32] An alternative to using binary variables would be to exploit the nuanced information within the responses by using Likert scales for each dimension. However, this approach does not improve the quality of the fit and introduces the challenge that the distribution of the Likert transformed variables is not normally distributed.

3. NTM (*there are Not Too Many immigrants*) takes a value of one when the respondent disagrees to some extent or strongly with statement S3; otherwise, it is equal to zero.
4. NLL (*Not all immigrants should Learn Luxembourgish*) takes a value of one when the respondent disagrees to some extent or strongly with statement S4; otherwise, it is equal to zero.

Figure 3.2.2 breaks down average pro-immigration attitudes by respondents' immigrant background. Overall, 77% and 76% of native-born respondents agree that immigration enriches the national identity and has a positive impact on the economy, respectively. While these percentages are similar to the national averages, pro-immigration attitudes are more pronounced among those with an immigrant background. This tendency is more pronounced among second-generation immigrants and respondents born in Portugal, neighboring countries (i.e., Germany, France, and Belgium), and non-European countries.

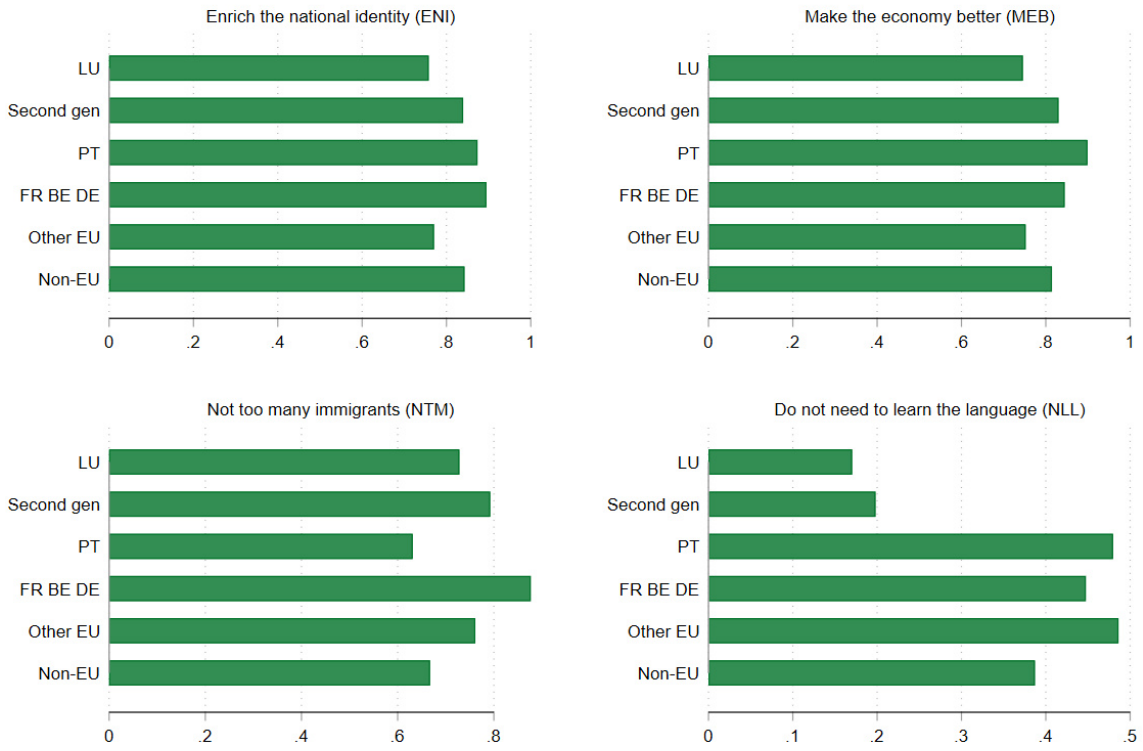


Figure 3.2.2: Pro-immigration attitudes: Differences across groups

Notes: Survey responses by respondents' immigrant background: those born in Luxembourg ("LU"), those born in Luxembourg to immigrant parents ("Second gen"), those born in Portugal ("PT"), bordering countries France, Belgium or Germany ("FR BE DE"), other EU- and non-EU countries.

Regarding perceptions of the optimal level of immigration, a majority of natives say that the current level of immigration is appropriate. Interestingly, this proportion varies between groups. It is generally higher among immigrants, with the exception of those from Portugal and non-European countries. Second-generation immigrants and respondents born in European countries show more favorable attitudes toward immigration levels. In contrast, a small proportion of native-born citizens, including

second-generation immigrants, believe that immigrants do not need to learn Luxembourgish. First-generation immigrants, in particular from EU countries, demonstrate a higher disregard for the necessity of linguistic integration.

Of course, these differences across groups could potentially be explained by differences in individual characteristics such as age, education, gender or labor market status. These effects will be examined in detail in the next empirical section. Finally, in Appendix 3.D, we develop a pro-immigration index (labeled *Pro-immig*) that summarizes the immigration views expressed in several questions of the survey. Perceptions of the pro-immigration index are consistent with those of economic and cultural perceptions of immigration. In addition, we show that the empirical analysis of the determinants of the *Pro-immig* index mirrors that of the ENI and MEB variables.

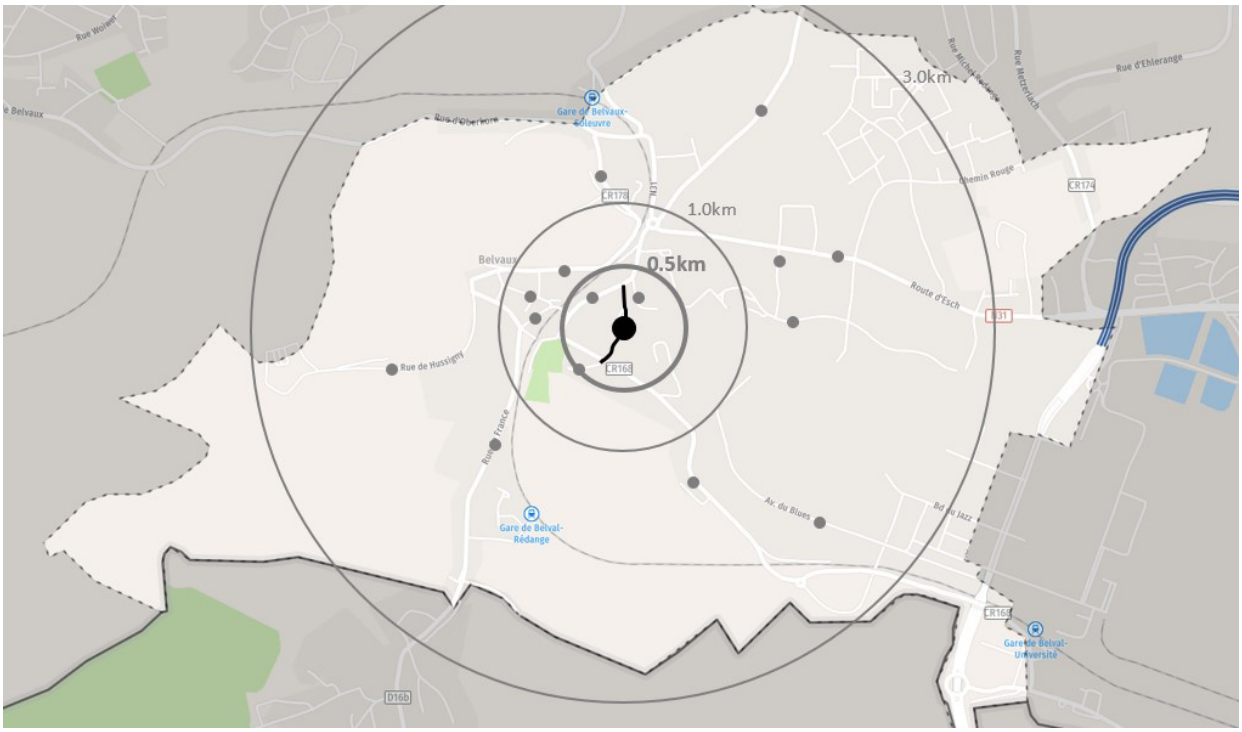
Neighborhood composition. Access to social security data enables us to accurately identify the respondent's place of residence at the zip code level, which can correspond to one street in Luxembourg and sometimes a subset of residences within a street. This information is available for the entire population of Luxembourg. As a result, we can determine the composition of the population in terms of birthplace in the immediate vicinity of the 1,954 respondents that consented to linking their data.

To achieve this, we geo-localize the centroid of each zip code in Luxembourg and use different buffer sizes to define the close neighborhood. By employing different buffer sizes, we define various levels of neighborhood proximity. The strictest definition considers only people living within the same zip code of residence. The more expansive definitions categorize individuals residing within a group of zip codes, the centroids of which are situated within a radius of 0.5, 1, or 3 kilometers from the centroid of the respondent's zip code of residence.

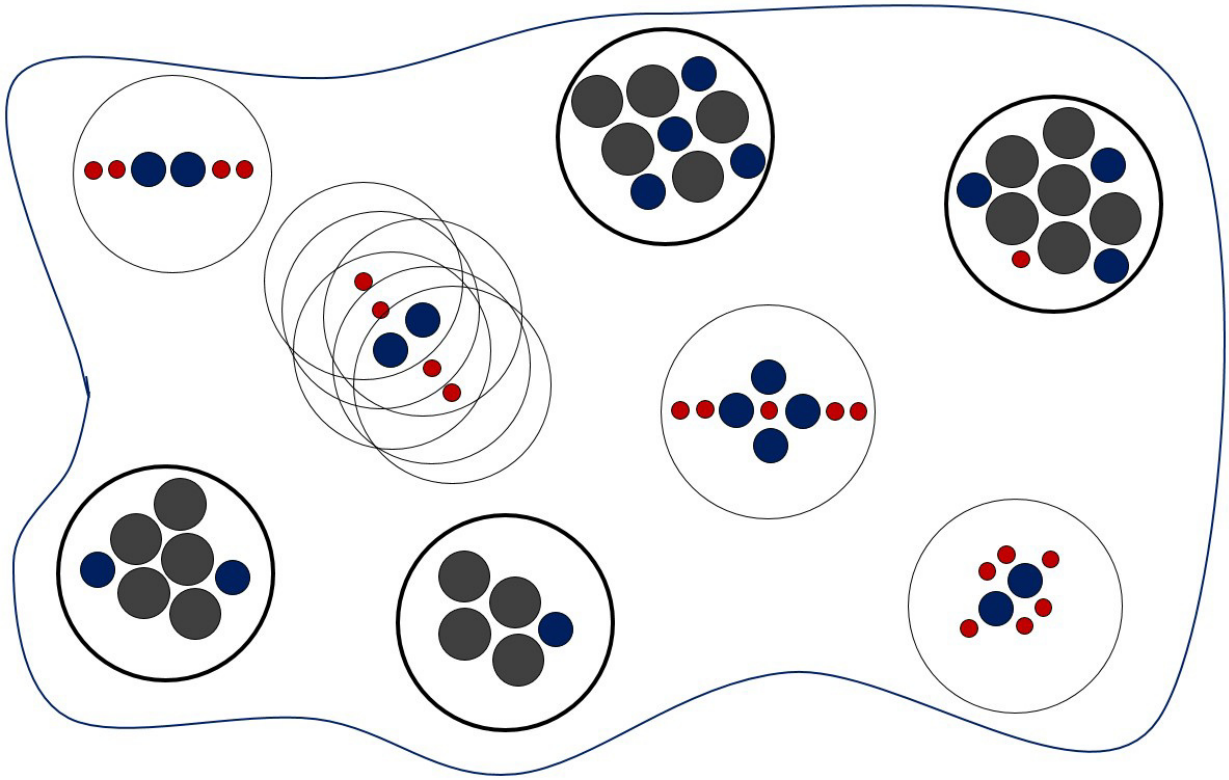
Naturally, individuals may also interact with people from diverse backgrounds at their workplace. Administrative social security data includes information about the employer of each employed respondent. Therefore, we can also analyze the workforce's composition in the respondent's workplace. As it is challenging to find an appropriate instrument for this variable, this part of our analysis is presented in the appendix.

Figure 3.2.3 visually illustrates our approach. Panel (a) portrays the street layout within the *Belvaux* municipality, situated in the southern region of Luxembourg. The street named *Rue du Stade* is highlighted in black, spanning a length of approximately 400 meters, with its centroid indicated by the black dot. This framework enables us to compute the proportion of migrants residing within that street (or zip code), corresponding to the immediate vicinity of the respondent's location. We represent this ratio as $\frac{M_{z,r,2020}}{P_{z,r,2020}}$. Here, $M_{z,r,2020}$ stands for the migrant population in the specific zip-code z within geographical area r in the year 2020, while $P_{z,r,2020}$ corresponds to the overall population in that same zip-code. This proportion can be further disaggregated across distinct migrant categories, delineated by their region of origin (including Europeans, non-Europeans, Portuguese individuals, and those born in neighboring countries) or by their year of arrival (encompassing the total immigrant population and those who have arrived within the last ten years). Specifically, the fraction $\frac{M^o_{z,r,2020}}{P_{z,r,2020}}$ signifies the relative presence of migrants from the origin country o within the population of the given neighborhood z.

Alternative buffer sizes are represented by the gray circles surrounding the centroid, each with radius of 0.5, 1, and 3 kilometers. Different metrics reflecting neighborhood composition can be derived by aggregating the populations residing within all zip codes



(a) Example of zip-code in *Belvaux*



(b) Hypothetical representation of immigrant share by buffer

Figure 3.2.3: Computation of the neighborhood composition

whose centroids (depicted as the gray dots) fall within a specific radius. This approach enables us to examine whether exposure to a particular type of immigrant affects attitudes towards immigration. The size of the buffer not only determines the pool of

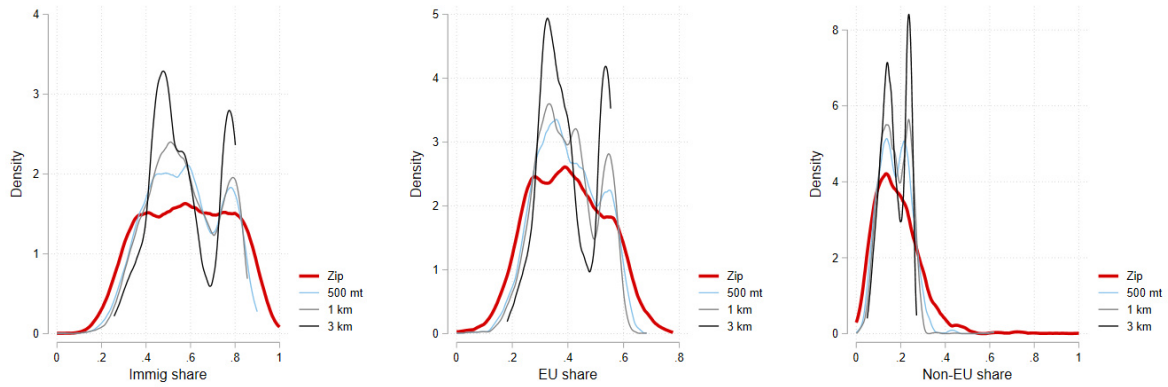
neighbors with whom the respondent is likely to interact but also affects the exogeneity of the neighborhood's composition, as discussed in the next section.

Modifying the buffer size induces alterations in the distribution of neighborhood composition and exposure variables. To elucidate this phenomenon, Panel (b) presents a hypothetical scenario involving a region comprising 60 distinct zip codes. Each zip code is depicted as a bubble, with its size and color denoting the proportion of immigrants in the respective population. Three categories of zip codes are delineated: small-red bubbles denote areas with a low immigrant population share, large-gray bubbles signify regions with a substantial immigrant presence, and medium-blue bubbles represent areas with a moderate immigrant share, falling between the low and high extremes. The allocation ensures an equal distribution of 20 bubbles in each category, resulting in a perfectly uniform distribution of immigrant shares across zip codes. The zip codes are further organized into eight distinct clusters, with no overlap between them.

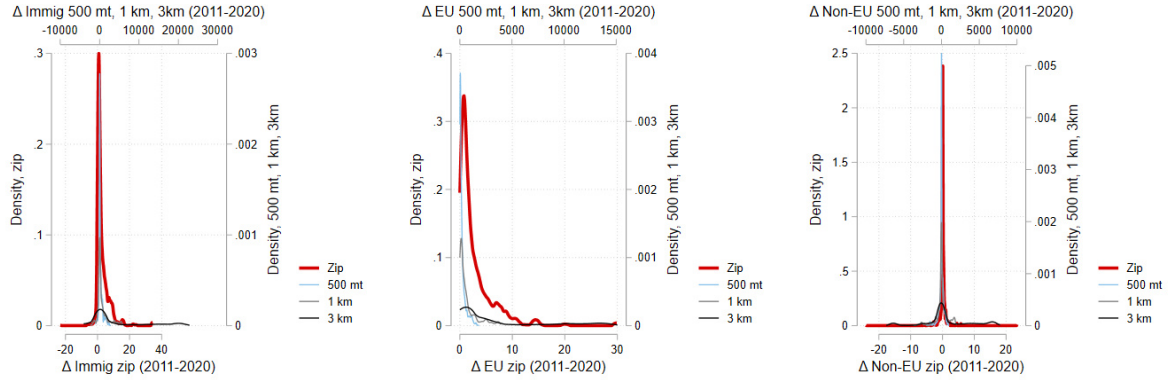
Consider buffers with a radius of 3 kilometers surrounding each centroid. The representation in Panel (b) simplifies the visualization by featuring a single large circle per buffer, with an exception in one buffer that better reflects our approach – a buffer is centered on each zip-code centroid. The uniform distribution of immigrant share across buffers is maintained when zip codes of varying sizes are randomly assigned to each buffer. In contrast, a polarized distribution emerges when large-gray and small-red bubbles seldom coexist in the same buffers, while medium-blue bubbles are randomly associated with both. In the presented example, this leads to the formation of four buffers with lower-middle immigrant shares and four with upper-middle shares. Hence, the distribution becomes bimodal when the buffer size increases. Conversely, an alternative configuration would be obtained if the distribution of immigrant shares by zip code is bimodal, and all buffers encompass both small-red and large-gray zip codes. In this case, increasing the buffer size results in a more uniform distribution of immigrant shares across the buffers.

Figure 3.2.4 depicts the distribution of the local exposure to immigration in Luxembourg. The upper panel shows the density the proportion of immigrants residing within the zip code or in streets whose centroids fall within a radius of 0.5, 1 or 3 kilometers in the year 2020. The density computed at the zip code level is rather uniform (thick red curve), and ranges between 30 and 85% when considering total immigration, between 25 and 65% for European immigration, and between 5 and 20% for non-European immigrants. When the radius increases, the distribution becomes more bimodal. For example, when considering the total share of immigrants living within a radius of 3 kilometers (thick gray curve), two peaks appear: one at 45% (which is slightly smaller than the nation-wide average share of foreign-born) and another one at 80%. Aligned with the hypothetical illustration in Panel (b) of Figure 3.2.3, this bimodal distribution underscores the geographic concentration of zip codes with lower-middle and upper-middle immigrant shares within distinct clusters. In an intuitive context, the second peak is indicative of numerous zip codes situated in the two principal cities, Luxembourg City and Esch-sur-Alzette, whereas the lower peak includes sparsely populated regions in the East and North of the country (Docquier et al., 2023).

Exposure to recent immigration. We are also interested in examining the effect of exposure to recent immigration waves in the close neighborhood, particularly in the last decade. Unfortunately, the social security data for earlier years are not sufficiently compatible with those of 2020, as the records were less precise during this period,



(a) Share of all immigrants in population (based on total stocks in 2020)



(b) Change in the count of immigrants (based on predicted variations between 2011 and 2020)

Figure 3.2.4: Distribution of the neighborhood composition measures

particularly when it comes to account for de-registrations from some municipalities.

To address this issue, we use statistics on the country-level population by citizenship in 2011 from the National Institute of Statistics and Economics (STATEC) to construct a proxy for local-level origin-specific stocks of immigrants in 2011. It is important to note that (i) these stocks encompass individuals, particularly young people, who may not be captured in social security records, (ii) we have to rely on citizenship rather than birthplace information, and (iii) our imputation formula operates under the assumption that the growth rates of the immigrant population are homogeneous across zip codes. Specifically, let g^o denote the national growth rate of the immigrant population from origin o over the last decade (i.e., between 2011 and 2020). Our estimate of the origin-specific migrant stock in 2011 at the zip code level is given by $\hat{M}^o_{z,r,2011} \equiv \frac{M^o_{z,r,2020}}{1+g^o}$. Our proxy for recent immigration flows from a given set O of countries of origin at the zip code level can be expressed as follows:

$$\Delta M^O_{z,r,2011-20} \approx M^O_{z,r,2020} - \hat{M}^O_{z,r,2011} = \frac{g^o M^o_{z,r,2020}}{1+g^o}. \tag{3.1}$$

The bottom panel of Figure 3.2.4 shows the density the predicted 2011-2020 change in the total number of immigrants within the zip code or in streets whose centroids fall within a radius of 0.5, 1 or 3 kilometers. As we are utilizing a count variable here, the $\Delta M^{O}_{z,r,2011-20}$ level inherently increases with the size of the buffer. The peak of the density is around 0 and the distribution is right-skewed, reflecting the increase in exposure to immigration in most locations.

3.3 Empirical strategy

We examine the factors influencing attitudes towards immigration in Luxembourg, with a specific focus on the impact of local exposure to immigration. Our general specification can be expressed as:

$$\text{Attitudes}_{i,z,r} = \alpha_r + \mathbf{X'}_{i,z,r}\beta + \mathbf{Z'}_{z,r}\gamma + \theta\text{Immig}^{O}_{z,r} + \epsilon_{i,z,r}, \tag{3.2}$$

where $\text{Attitudes}_{i,z,r}$ is a dummy variable equal to one if individual i living in the zip code (or buffer area) z of the administrative location r (i.e. municipality or districts)[33] has a positive attitude towards immigration, as defined by the indicators *ENI* (enriches identity), *MEB* (enriches economy), *NTM* (not too many immigrants), and *NLL* (should not learn Luxembourgish). $\mathbf{X}_{i,z,r}$ is a set of individual covariates which includes the foreign background and other demographic characteristics, $\mathbf{Z}_{z,r}$ is a set of zip-code or buffer level average characteristics, α_r is a fixed effect that absorbs potential confounding heterogeneity at the level of the administrative location r, and $\epsilon_{i,z,r}$ is the error term. Alternatively, $\text{Immig}^{O}_{z,r}$ can quantify either the proportion of immigrants from a designated group of origin nations O residing in neighborhood z, represented as $\frac{Mz,r,2020^{O}}{P^{O}_{z,r,2020}}$, or the count of recent immigrants from O, denoted as $\Delta M^{O}_{z,r,2020}$. [34]

We begin our analysis by using a linear probability model to estimate Eq. (3.2). However, a common issue in the literature on neighborhood effects is the potential for endogeneity in the composition of the neighborhood. This is because both migrants and natives choose where they live to optimize their well-being. For example, migrants may prefer areas where the population is more welcoming towards immigrants. In Luxembourg, where housing is scarce and expensive, finding suitable accommodations can be a daunting task for newcomers. While individuals may be able to choose the canton or municipality in which they live, selecting an exact residence is often challenging, if not impossible. To mitigate endogeneity concerns, we restrict our main neighborhood definition to the zip code (street) of residence, where variations in population composition are more likely to be quasi-random for a given administrative

[33] The index r denotes the 102 municipalities in Luxembourg, which serve as the smallest administrative units within the country. For the two largest municipalities, namely Luxembourg and Esch-Sur-Alzette, r undergoes further disaggregation to signify distinct city districts. To address potential concerns regarding the granularity of location information and the potential breach of respondent anonymity, we are permitted to observe pseudonymous identifiers for both r and z. These pseudonymous identifiers facilitate the computation of origin-specific population counts by zip code, as well as a range of zip code-specific characteristics detailed in Appendix 3.4.1. Furthermore, these identifiers allow us to associate each zip code with its corresponding municipality or district, enabling the incorporation of the relevant fixed effect.

[34] In Appendix 3.E, we consider a similar specification but focus on the role of the birthplace composition of the work force of their main employer. For this regression analysis, our sample is limited to respondents who are employed.

unit (Bayer et al., 2008). Controlling for locality level fixed effects, we exploit the variation between zip codes within administrative localities (r).

It can still be argued that people with similar characteristics decide to live in close proximity (Goux & Maurin, 2007; Manski, 1993). This becomes especially evident when we broaden the neighborhood definition to encompass a larger area. While this expansion might yield more accurate measurements of local immigration exposure, it also intensifies concerns related to endogeneity. To address these concerns, some studies use close neighborhood characteristics to study individual attitudes or social/economic outcomes, but also use additional sources of exogenous variation such as the allocation of public housing or resettlement programs as instrumental variables (Algan et al., 2016; Katz et al., 2001; Kling et al., 2007; Oreopoulos, 2003). In this study, we use an instrumental variable approach that relies on Bartik-type instruments (Bartik, 1991), even when limiting the buffer size to the zip code of residence.

Our instrumental approach involves two steps. First, we allocate the total stock of immigrants from each origin o across zip codes based on their shares at an initial period. For this, we use social security data from 2010.[35] The predicted number of migrants from origin o, living in the zip code area z of locality r in 2020 is given by:

$$\widehat{M}^{o}_{z,r,2020} = \frac{M^{o}_{z,r,2010}}{\sum_{z',r'} M^{o}_{z',r',2010}} \times \sum_{z',r'} M^{o}_{z',r',2020}, \tag{3.3}$$

where the second term signifies the overall immigrant population from the origin country o during the survey year 2020 (referred to as the "shift" component), and the first term encapsulates the dispersion of these immigrants across zip codes z in the year 2010 (termed the "share" component).

In the second step, we aggregate these origin-specific stocks over groups/regions O and divide by the total population in area z, $P_{z,r,2020}$, to construct an instrument for $\text{Immig}^{O}_{z,r}$:

$$\widehat{Immig}^{O}_{z,r} = \frac{\sum_{o \in O} \widehat{M}^{o}_{z,r,2020}}{P_{z,r,2020}}. \tag{3.4}$$

We note that the overall population figure might also be susceptible to endogeneity for a couple of reasons. Firstly, it encompasses the migrant population itself, and secondly, local residents might choose to relocate to regions with lower migrant proportions. Nevertheless, within a constrained housing market like Luxembourg's, where virtually all housing units are occupied, the total population effectively mirrors the available housing stocks in the zip codes or buffer zones denoted by z. In addition, our survey data show that very few of the natives express a preference to avoid a neighbor of a particular origin or with a foreign background.[36] However, we recognize that we are unable to rigorously assess the strength of the exogeneity assumption on the initial shares as suggested by Goldsmith-Pinkham et al. (2020), given that we do not have any

[35] While this data comes with compatibility issues with respect to the 2020 data as mentioned above, for the instrument we refrain from using more reliable STATEC data as it would involve a double-prediction process, which makes it difficult to recover proper standard errors. Utilizing the observed allocation data from 2002, the earliest year accessible in the social security database, proves effective in instrumenting the current allocation of European immigrants. However, the 2002 shares exhibit weak predictive capability for the allocation of non-European immigrants, particularly those who have primarily arrived in recent years. Thus, the 2010 data would be the best available choice.

[36] We use this information in a robustness check in order to assess the relevance of this "native flight" channel.

time variation in our outcome variable.[37] Our shift-share approach therefore mainly serves as a robustness check by an alternative measure of immigration exposure.

In our regression of attitudes on changes in immigration ($\Delta M^{O}_{z,r,2020}$), we employ the predicted variations in immigration derived from Eq. (3.1), akin to a shift-share approach. It is thus essential to consider the added uncertainty inherent in these predictions, as it could introduce bias to the estimated standard errors of our "generated regressors" of interest (Wooldridge, 2002). Given that assessing the prediction's accuracy is unfeasible, rectifying the variance-covariance matrix analytically is not possible. To address this, we utilize a bootstrapping technique, involving repetitive re-sampling of the data with replacement across an extensive number of iterations. This procedure, with the sampling variation, generates a distribution of the standard errors and allows to compute bootstrapped estimates.

3.4 Results

This section presents the main findings of our empirical research. First, we use OLS regressions to examine the importance of socio-demographic factors in shaping attitudes toward immigration, with a particular focus on respondents' foreign background. Although our dependent variable is binary, we opt for a linear probability model (rather than a non-linear model based on a logit or probit) which allows a simple and direct interpretation of the estimated coefficients as well as a comparison between our alternative models. In practice, the linear probability model and the non-linear models tend to produce similar marginal effects when the values of the independent variables are close to their mean values (Wooldridge, 2013; Wooldridge & Pischke, 2008).

We show that the foreign background of the respondents is a key determinant of these attitudes. We then turn our attention to assessing the influence of neighborhood composition through the implementation of OLS and 2SLS regressions. We find that exposure to immigrants in the immediate neighborhood has a limited effect on attitudes, with the exception of recent inflows changing natives' perceptions of the optimal level of immigration. Opinions on benefits related to the economy or national identity are not significantly affected. These results are not specific to any particular subgroup of native-born respondents, as shown by the extensive heterogeneous analysis in Appendix 3.C.3.

The Influence of Migration Background (OLS). – Table 3.4.1 presents the results of OLS regressions in which the four dependent variables are regressed on individual characteristics including gender, age, education, marital and occupational status, religiosity (top panel) and immigrant background (bottom panel). The reference category includes women born in Luxembourg, aged 35 to 55, with an upper secondary education and a non-practicing religious attitude (i.e., religiosity = 0). In terms of migration background, we distinguish between second-generation immigrants (born to at least one foreign-born parent) and first-generation immigrants from Portugal, neighboring countries (namely France, Germany, or Belgium), other European Union (EU) member states, or non-European countries (non-EU).

[37] For example, Goldsmith-Pinkham et al. (2020) outline potential diagnostics including testing correlation with observables that predict *changes* in the outcome, and checking pre-trends in the outcome across regions with different immigration shares.

Table 3.4.1: Effect of socio-demographic characteristics and migration background (OLS)

	ENI	MEB	NTM	NLL
Male	0.027	0.042**	0.047**	0.025
	(0.018)	(0.019)	(0.021)	(0.022)
Below35	-0.025	-0.033	-0.056**	0.017
	(0.023)	(0.024)	(0.027)	(0.029)
Above55	-0.004	0.006	-0.021	-0.030
	(0.024)	(0.026)	(0.030)	(0.031)
Low educ.	-0.140***	-0.092**	-0.057	0.051
	(0.036)	(0.038)	(0.043)	(0.045)
College educ.	0.063***	0.025	0.082***	0.039
	(0.021)	(0.023)	(0.026)	(0.027)
Unemployed	-0.050	-0.070*	-0.094**	-0.095**
	(0.038)	(0.041)	(0.046)	(0.048)
Farmer or blue collar	-0.036	-0.061*	-0.006	-0.088**
	(0.032)	(0.035)	(0.039)	(0.041)
Married	0.007	0.007	0.000	-0.018
	(0.020)	(0.021)	(0.024)	(0.025)
Children Below 26	0.007	-0.026	0.004	-0.001
	(0.020)	(0.022)	(0.025)	(0.025)
Religiosity	-0.005	-0.004	-0.095***	-0.081***
	(0.019)	(0.020)	(0.023)	(0.024)
Second gen immigrant	0.087**	0.104**	0.096*	0.015
	(0.043)	(0.046)	(0.052)	(0.054)
FR-BE-DE	0.118***	0.085**	0.105***	0.211***
	(0.033)	(0.036)	(0.040)	(0.042)
PT	0.178***	0.225***	-0.012	0.343***
	(0.037)	(0.039)	(0.044)	(0.046)
OtherEU	0.001	0.022	0.002	0.240***
	(0.043)	(0.046)	(0.052)	(0.054)
NonEU	0.077***	0.087***	-0.065**	0.186***
	(0.026)	(0.028)	(0.031)	(0.033)
Observations	1876	1876	1876	1876

Notes: *** p<0.01, ** p<0.05, * p<0.10. These OLS regressions show the estimates of the correlation of individual characteristics on opinions toward immigration as captured by the dependent variables *ENI* (enriches identity), *MEB* (enriches economy), *NTM* (not too many immigrants), and *NLL* (should not learn Luxembourgish). The dependent variables are defined as dummies that capture pro-immigration attitudes towards immigration. The regressions control for a set of zip-code characteristics (immigration share and predicted change in number of immigrants between 2011 and 2020, log median income, log share of individuals receiving social benefits, log of population), as shown in Table 3.C.1. All regressions are based on the full sample (Luxembourg- and foreign-born) and include a locality fixed effect. Standard errors are clustered at the locality level.

All regressions are based on the full dataset and include a locality fixed effect coupled with a set of controls measured at the zip code level (these effects will be discussed later): the share of foreign-born residents in 2020, the predicted change in the foreign-born population between 2011 and 2020, the log-transformed level of median income, the share of individuals receiving welfare, and population size. We will focus on these variables in the next section.

The top panel of the table shows that pro-immigration sentiments tend to be more pronounced among male respondents, with significant associations with attitudes related to the economy and the level of immigration. Education also emerges as an influential factor: compared to the reference groups of secondary school graduates, respondents with lower levels of education demonstrate significantly lower perceptions of the benefits of immigration to the national identity and the economy, while tertiary education is positively correlated with the perceived level of immigration. Age, marital status and having children, on the other hand, do not seem to be significantly correlated with opinions, except for those aged 35 and under, who are more likely to believe there are too many immigrants in Luxembourg. Those who are unemployed or employed in low-skilled occupations also show reduced pro-immigration attitudes. Religiosity emerges as a notable factor, particularly associated with lower pro-immigrant and more assimilationist views. In Appendix 3.C.3, we show that these effects are sometimes more pronounced among natives.

More importantly, the bottom panel of the table highlights how attitudes towards immigration are correlated with respondents' own foreign background. Compared to individuals with no foreign background, being a second-generation immigrant is positively related to perceived benefits of immigration: the probability of viewing immigrants as culturally and economically beneficial increases by 8.7 to 10.4 percentage points (henceforth pp). Given that the average likelihood of these positive attitudes for native-born respondents is about 75% (as shown in table 3.B.1), these effects are important. This group is also more likely (by 9.6 pp) to say that the number of immigrants is appropriate, which is noteworthy given that the average probability is 54.5%. However, there is no difference between second-generation immigrants and other natives when asked whether immigrants should be required to learn Luxembourgish.

These effects are even more pronounced among first-generation immigrants. Respondents born in Portugal are 17.8 to 22.5 pp more likely than the native reference group to believe that immigration is culturally and economically beneficial. This difference corresponds to 11.8 and 8.5 pp for respondents born in bordering countries (denoted by FR-BE-DE), and 7.7 and 8.7 pp for those born in non-European countries. Interestingly, among respondents born in other European Union Member States, the difference with the native born is not statistically significant, except for their attitude towards immigrants' need to learn Luxembourgish.

Regarding opinions on the optimal level of immigration, the probability of considering the current level of immigration to be appropriate is higher among respondents from bordering countries. However, there is no significant difference between natives and those born in Portugal or in other (non-bordering) EU Member States, while being born in a non-EU country reduces this probability by 6.5 percentage points.

Finally, all first-generation immigrants are more likely to think that immigrants should not be obliged to learn Luxembourgish. The largest difference (34.3 pp) is observed among respondents born in Portugal, followed by those from other European Union countries (24.0 pp), bordering countries (21.1 pp) and the rest of the world (18.6 pp). It should be noted that these differences are substantial, given that the average

probability is 29.3% within the full sample.

In sum, our findings reveal a strong correlation between attitudes toward immigration and sociodemographic characteristics, consistent with previous research by Drazanova et al. (2022), Facchini and Mayda (2017) and Hainmueller and Hopkins (2014). In particular, we emphasize that the respondent's foreign background emerges as a key determinant of these attitudes, extending the contributions of Bursztyn et al. (2024) and Gihleb et al. (2022). In particular, being born in Portugal and neighboring countries increases pro-immigration sentiments among residents.

The Impact of Local Exposure to Immigration (OLS and 2SLS). – Table 3.4.2 shows the results of OLS and IV regressions where our four different dependent variables are regressed on the share of immigrants at the zip code level. Columns (1)-(4) use the full sample of respondents (as in Table 3.4.1), columns (5)-(8) restrict the sample to Luxembourgish respondents, and columns (9)-(12) restrict the sample to foreign-born respondents. Panel (a) presents the OLS results for the total immigrant share, while in Panel (b) we focus on recent immigrants who arrived within the last decade. In Panel (c), we return to the overall immigrant share and adopt the shift-share IV methodology outlined in Section 3.3. In all of these regressions, we control for local fixed effects, individual characteristics including immigrant background, and the range of zip code characteristics detailed in Table 3.4.1.

Analysis of Panel (a) reveals that the association between the share of immigrants in the neighborhood and pro-immigration sentiments is negative in most specifications (at least in the full and native samples), albeit statistically insignificant, regardless of the dependent variables and the population under study. This result is confirmed in the 2SLS regression in Panel (c). In Panel (b), we shift our focus to recent immigrants, in particular, those who arrived in the last decade. Standard errors are computed using bootstrapping techniques. An increase of 10 new immigrants in the zip code leads to a substantial 27 percentage point decrease in the probability of perceiving the number of immigrants as reasonable (NTM).[38] Caution is needed in interpreting these results, as the number of recent immigrants is a constructed proxy. However, given the fact that the average probability equals 54.5%, these results suggest that exposure to recent immigration may have a pronounced effect on this dependent variable compared to the overall share of immigrants.

We infer that recent immigration inflows in the immediate neighborhood may contribute to increased concerns among natives about the optimal level of immigration. However, these effects do not manifest themselves in immigrants' perceptions (last four columns). Moreover, neither the share of immigrants nor the number of recent immigrants is significantly associated with other pro-immigration attitudes. This includes perceptions of the cultural and economic benefits of immigration as well as views regarding learning the local language.

In Appendix 3.C.3, we delve into the full sample and introduce an interaction term between the zip code-level immigrant share and certain individual characteristics, including advanced age, education, economic inactivity, and foreign background. Most of these interactions are not statistically significant at the 5% level. The negative effect of immigration on the opinion about immigration levels is slightly larger among older

[38] Without boostrapping, the standard error is 0.010 and the estimate is significant at the 1% threshold. When the number of boostrap replications is greater than 100, the standard error reaches 0.012 and the coefficient remains significant at the 5% level. This result is maintained when the number of bootstrap replications is further increased, as shown in Appendix 3.C.4.

Table 3.4.2: Effects of neighborhood exposure to immigration

	Full sample				Native sample				Immigrant sample			
	(1) ENI	(2) MEB	(3) NTM	(4) NLL	(5) ENI	(6) MEB	(7) NTM	(8) NLL	(9) ENI	(10) MEB	(11) NTM	(12) NLL
Panel a												
Immig share zip (OLS)	-0.042	-0.044	-0.120	0.079	-0.097	-0.085	-0.234	0.090	0.020	0.029	-0.033	0.104
	(0.065)	(0.091)	(0.113)	(0.090)	(0.158)	(0.158)	(0.248)	(0.185)	(0.080)	(0.106)	(0.115)	(0.117)
Observations	1876	1876	1876	1876	487	487	487	487	1389	1389	1389	1389
Panel b												
Δ Immig (2011-2020)	-0.005	-0.004	-0.000	0.002	-0.005	0.012	**-0.027****	-0.001	-0.004	-0.006	0.006	0.003
	(0.003)	(0.003)	(0.004)	(0.004)	(0.012)	(0.011)	(0.012)	(0.015)	(0.004)	(0.004)	(0.004)	(0.005)
Observations	1876	1876	1876	1876	487	487	487	487	1389	1389	1389	1389
Panel c												
Imm share zip (2SLS)	-0.035	-0.149	-0.406	0.407	-0.101	-0.032	-0.745	0.233	-0.019	-0.234	-0.293	0.578
	(0.312)	(0.287)	(0.435)	(0.387)	(0.425)	(0.432)	(0.515)	(0.491)	(0.461)	(0.426)	(0.631)	(0.588)
Observations	1797	1797	1797	1797	452	452	452	452	1314	1314	1314	1314
Kleibergen-Paap Wald rk F statistic	35.945	35.945	35.945	35.945	17.040	17.040	17.040	17.040	20.544	20.544	20.544	20.544
Cragg-Donald Wald F statistic	129.807	129.807	129.807	129.807	65.932	65.932	65.932	65.932	68.373	68.373	68.373	68.373
Controls (as in Table 3.4.1)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Locality FE (r)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Lux	Lux	Lux	Lux	Imm	Imm	Imm	Imm

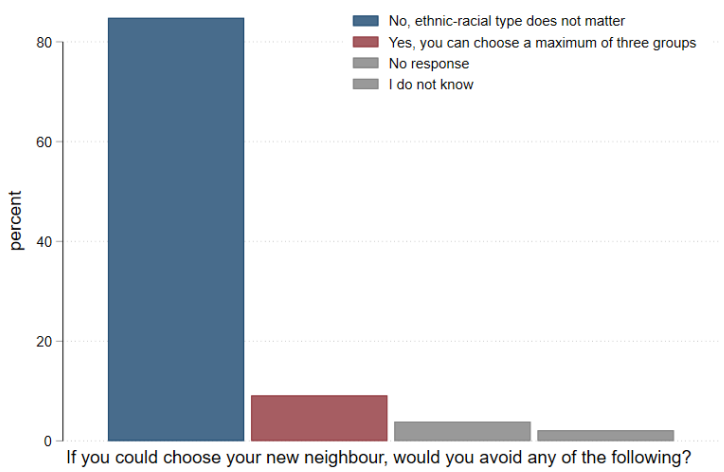
Notes: *** p<0.01, ** p<0.05, * p<0.10. The regressions show the estimates of the effect of immigration on opinions toward immigration as captured by the dependent variables *ENI* (enriches identity), *MEB* (enriches economy), *NTM* (not too many immigrants), and *NLL* (should not learn Luxembourgish). The dependent variables are defined as dummies that capture pro-immigration attitudes towards immigration. Columns (1)-(4) use the full sample of respondents, columns (5)-(8) restrict the sample to Luxembourgish respondents, and columns (9)-(12) restrict the sample to foreign-born respondents. Immig share zip is the proportion of immigrants living in a zip-code, relative to the total zip-code population. All regressions control for a set of individual-level characteristics (gender, age, education level, being married, having children, blue collar or farmer, being religious, being unemployed, migration background and when the sample includes the foreign-born, origin dummies) as well as a set of zip-code characteristics (log median income, log share of individuals receiving social benefits, log of population). All regressions include a locality fixed effect. In Panel (a) we estimate OLS regressions and we report robust standard errors clustered at the locality level in parentheses. In Panel (b) we estimate OLS coefficients on the imputed changes in immigration between 2011 and 2020 and we report bootstrapped standard errors in parentheses. In Panel (c) we estimate 2SLS regressions, where Immig share is instrumented with a shift-share instrument, with the corresponding Kleibergen-Paap Wald rk F statistics and the Cragg-Donald Wald F statistics for weak instruments.

(aged 55 and over), low-skilled and inactive respondents, although these interactions are imprecisely estimated. In the same appendix (3.C.3), we show the results of a similar heterogeneous analysis on both the samples of native-born and immigrant. Remarkably, the results are identical. These results underline the robustness of our results presented in Table 3.4.2 across all respondent subgroups.

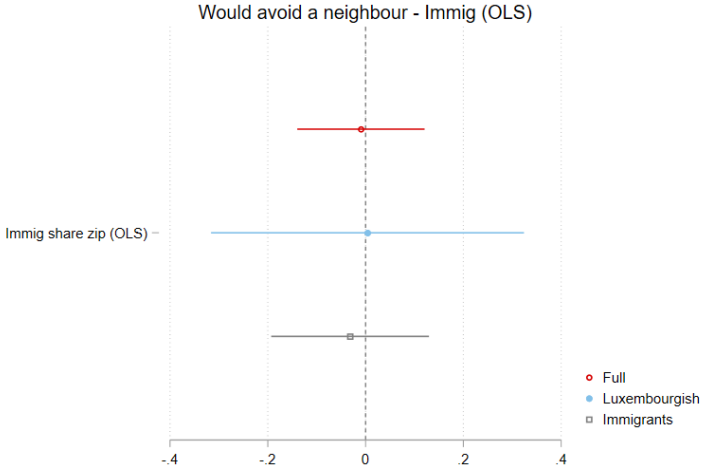
In Luxembourg, a country with a long history of immigration where migrants have become an integral part of the social fabric, our research shows that proximity to immigrants of different origins does not have a discernible effect on attitudes towards immigration. This might be interpreted as a lack of strong evidence for either the contact or the conflict theory at work. The absence of a negative effect on pro-immigration attitudes may also be related to the strong complementarities between immigrants and natives in the labor market (Verheyden et al., 2024). This is consistent with Bauer et al. (2000), who find that selection of immigrants based on labor-market needs encourages more positive attitudes towards immigration.

Native Flight Concerns. – A bias that could lead to an underestimation of the negative impact of local exposure to immigration on pro-immigration opinions is the "native flight" phenomenon (Alesina & Tabellini, 2022). This situation occurs when natives or other groups of residents who have negative attitudes towards recent immigrants decide to move to another neighborhood with fewer immigrants within Luxembourg, or even decide to reside abroad in the Greater Region and commute to work on a daily basis (Boesen, 2020). In our specific context, this phenomenon may be unlikely due to the lack of hostility towards immigrants – remember that a large majority of the population believes that immigration enriches the identity of the country and is beneficial to the economy – and due to the tightness of the housing market in Luxembourg.

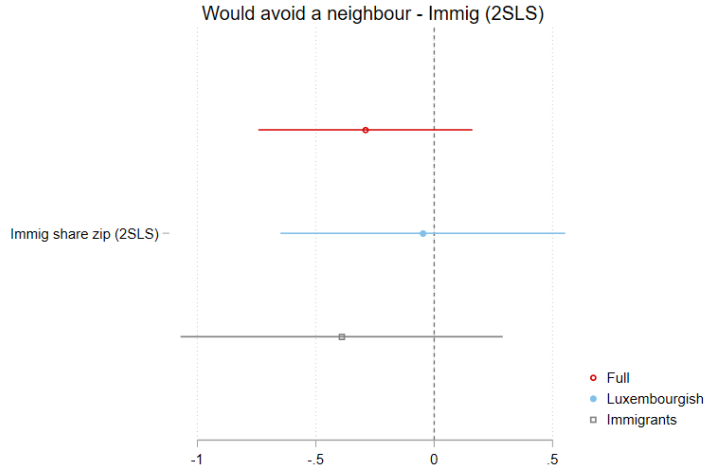
To more explicitly assess the plausibility of "native flight," we use an additional survey question asking whether respondents would prefer to avoid a neighbor who belongs to a particular ethnic minority group. In Panel (a) of Figure 3.4.1, we observe that less than 10% of respondents report such a preference. We also regress the likelihood of expressing a preference to avoid a neighbor of a particular group on our explanatory variables. Panel (b) shows 95% confidence intervals around the OLS estimates. The results show that, regardless of the sample under consideration, respondents who express a preference to avoid a neighbor of a particular group do not consistently reside in areas with higher immigration. Moving to Panel (c), the 2SLS estimates are plotted and shown to be not significant. Finally, in Panel (d), changes in immigration also have no significant effect on our constructed avoidance indicator.



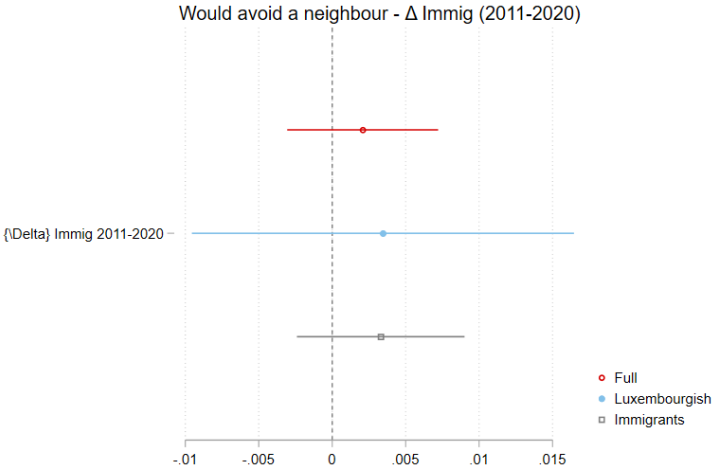
(a) Willingness to avoid a neighbor (as %)



(b) Effect of immigration share (OLS)



(c) Effect of immigration share (2SLS)



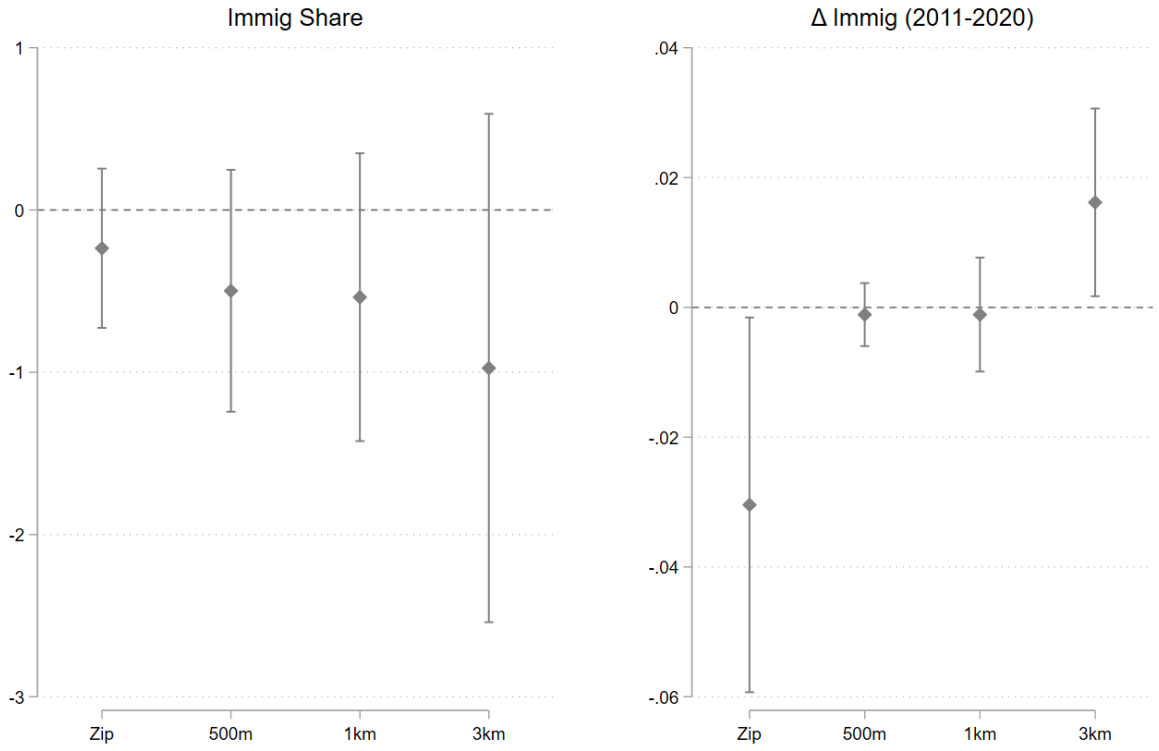
(d) Effect of recent migration inflow

Figure 3.4.1: Immigration and willingness to avoid a neighbor of a certain ethno-racial type

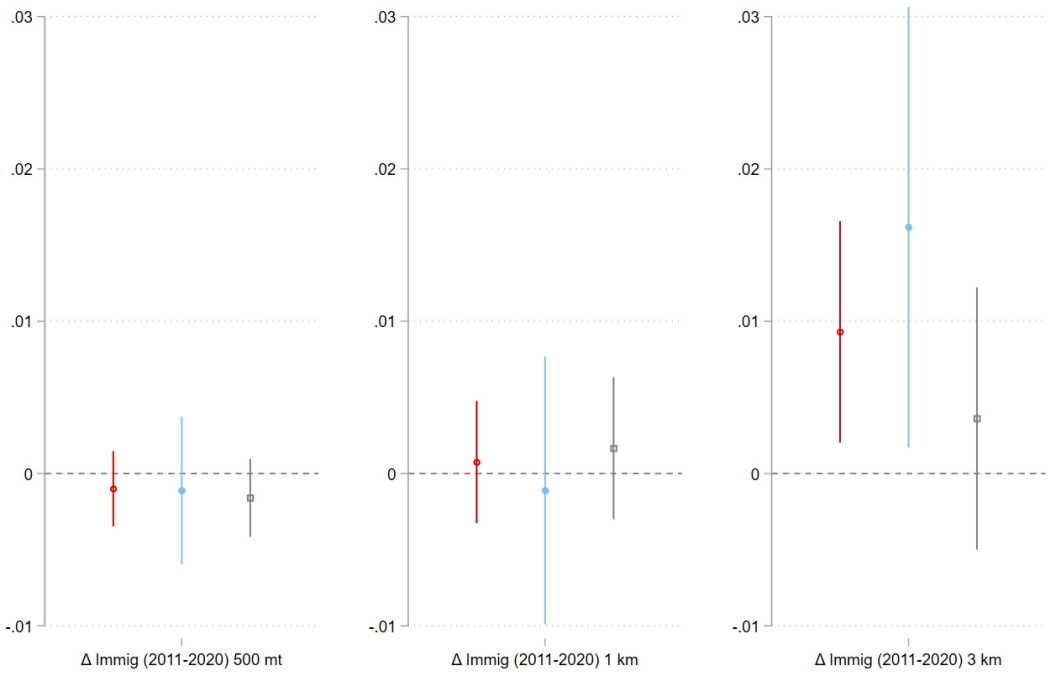
Robustness to the Neighborhood Definition. – Focusing on the opinion on the level of immigration (NTM) that shows significant associations with the recent inflows of migrants, we test whether this effect is influenced by the definition of the neighborhood. In the top panel of Figure 3.4.2, we focus on the effect of the total share of immigrants and the number of recent immigrants on the probability that native-born respondents consider the number of immigrants to be reasonable in the native sample. We standardize the immigration share or flow measured at different buffer levels by the number of zip codes included in each buffer. This ensures that all estimates are comparable, as they capture the effect of increasing the average immigration share or number of recent immigrants per zip code in the close or extended neighborhood.

We compare the effect obtained at the zip code level with that obtained when we include immigrants living in streets with a centroid within 0.5, 1, or 3 kilometers of the respondent's zip code. The insignificant effect of the total share of immigrants is maintained in all specifications. When focusing on the number of recent immigrants, the only negative and significant change emerges when restricting the variable of interest to the zip code (as seen before). The effect becomes positive and significant in the 3 km specification. On average, a zip code in Luxembourg corresponds to a street a few hundred meters long. In rural areas, adding zip codes with a centroid within 3 km can delineate an area of 12 to 20 km squared. Thus, although the result could be seen as confirming the contact theory (Allport, 1954), increasing the buffer size to 3 kilometers could introduce endogeneity because immigrants are more likely to choose broad locations where the incumbent population is more welcoming, which is practically impossible when focusing on a particular zip code.

The bottom panel of Figure 3.4.2 zooms in on the estimates obtained with a radius of 0.5, 1, and 3 kilometers and distinguishes between the effect obtained in the full, native, and immigrant samples. The positive effect in the 3 km specification is significant only in the native sample. Again, this result should be treated with caution, as immigrants are more likely to choose wide locations where natives are more welcoming. Therefore, we conclude that the results obtained from focusing on the respondent's immediate neighborhood are more convincing.



(a) Effects of immigration shares and counts in the native sample



(b) Zoom in on immigration counts in the full, native and immigrant samples

Figure 3.4.2: Robustness to the definition of the close neighborhood Effect of immigration on NTM

Heterogeneity along the Skill vs. Cultural Dimension. – An important insight that emerges from the existing literature on attitudes toward immigrants is that the effect tends to be driven by culturally distant and/or economically vulnerable immigrant subpopulations. To understand whether this is the case in our setting, we decompose the immigrant shares and counts along two dimensions in Table 3.4.3. In Panels (a) and (b), we distinguish between EU and non-EU immigrants as a proxy for cultural distance. EU immigrants are defined as immigrants born in a European Union member state, while non-EU immigrants are those born elsewhere. In Panels (c), we distinguish between low-income and high-income immigrants and compute their share in the population.[39] Low-income immigrants are defined as foreign-born individuals with income levels in the bottom quartile of the immigrant income distribution. High-income immigrants are those with income levels in the top quartile. These results can only be computed with OLS regressions because our Bartik instrument is not strong enough to predict the actual level of immigrant subpopulations. However, to mitigate endogeneity concerns, we control for the share of high- and low-income natives in the zip codes and for nativity dummies.

Panel (a) shows that the effect of the share of non-EU immigrants on the attitude towards the country's immigration level is weakly significant (at the 10% level) in the native-born sample, while that of the share of EU immigrants is insignificant. In Panel (b), the coefficient associated with the number of recent non-European immigrants is four times larger than that of recent European immigrants in the native-born sample, although the estimate is not statistically significant. No such level difference is observed in the foreign-born sample. Recall that the number of recent migrants is imputed using national growth rates of the immigrant population by origin over the past decade, combined with origin-specific shares by zip code from 2011 (see Eq. (3.1)). Because the size and origin mix of non-European immigration has changed dramatically over the past decades, the 2011 shares have less predictive power. Hence, the effects in Panels (a) and (b) are imprecisely estimated. However, they provide suggestive evidence that local exposure to non-European immigration may reduce natives' perceptions of the optimal level of immigration.

In Panel (c), the coefficient associated with the immigrant share is also negative and stronger for low-income immigrants, although it is insignificant in the full sample and when focusing only on the native-born. Turning to the sample of immigrant respondents, the share of low-income immigrants reduces the likelihood of having a pro-immigrant attitude, which is statistically significant at the 10% level. Our results suggest that exposure to these groups may influence perceptions of optimal immigration levels among earlier cohorts of immigrants (mostly Europeans), but not among natives.

[39] Without accurate data on the distribution of immigrants by income in 2011, we cannot proxy the number of recent immigrants by income quartile.

Table 3.4.3: Heterogeneity along the cultural vs. economic dimensions – Effect of neighborhood exposure on NTM (OLS estimates)

	(1) Full sample	(2) Native-born	(3) Foreign-born
Panel a			
EU share	-0.110	-0.052	-0.039
	(0.158)	(0.327)	(0.155)
Non-EU share	-0.132	-0.663*	-0.026
	(0.125)	(0.344)	(0.144)
Panel b			
Δ EU (2011-20)	0.003	-0.012	0.005
	(0.005)	(0.017)	(0.006)
Δ Non-EU (2011-20)	-0.008	-0.045	0.006
	(0.009)	(0.039)	(0.009)
Panel c			
Low-inc share	-0.579	-0.655	-0.668*
	(0.363)	(0.743)	(0.392)
High-inc share	-0.179	-0.254	-0.080
	(0.130)	(0.287)	(0.138)
Observations	1877	487	1390
Controls	Yes	Yes	Yes
FE	Yes	Yes	Yes

Notes: *** p<0.01, ** p<0.05, * p<0.10. The regressions show the estimates of the effect of immigration on opinions toward immigration as captured by the dependent variable *NTM* (not too many immigrants). The first column uses the full sample of respondents, column (2) restricts the sample to native respondents, the last column restricts the sample to foreign-born respondents. EU share is the proportion of immigrants born in the European Union and living in a zip-code, relative to the total zip-code population. Non-EU share is the the proportion of those born outside the European Union. Low-income immig is the proportion of immigrants belonging to the bottom quartile of the immigrants' income distribution over the total zip-code population, while High-income immig is the proportion of immigrants with a higher income than the bottom quartile of the immigrants' income distribution. Δ EU (2011-2020) and Δ Non-EU (2011-2020) are the predicted changes in the counts of EU and Non-EU immigrants by zip-code. The regressions control for a set of individual-level characteristics (gender, age, education level, being married, having children, blue collar or farmer, being religious, being unemployed, migration background and when the sample includes the foreign-born, origin dummies) as well as a set of zip-code characteristics (log median income, log share of individuals receiving social benefits, log of population). The regressions in Panel (c) include the share of low-income Luxembourg-born, therefore the reference category for Low-income immig and High-income immig is the proportion of native-born with a higher income than the first quartile of the immigrant income distribution. All regressions include a locality fixed effect. Standard errors are in parentheses. Panel (a) and Panel (c) report cluster robust standard errors. Panel (b) reports bootstrapped standard errors.

3.5 Conclusion

Combining survey data on attitudes toward immigration with social security records, we examine the determinants of pro-immigration sentiment in Luxembourg. With a substantial history of immigration and over 74% of the population with a migration background, Luxembourg provides an ideal setting for our investigation. Our sample includes native-born residents as well as first- and second-generation immigrants, giving us a comprehensive picture of attitudes toward immigration. Our focus is on the influence of foreign background and local exposure to immigrants, as measured by factors such as the proportion of immigrants or the number of recent immigrants in each respondent's immediate neighborhood.

Our study provides valuable insights into the complex interplay of factors that shape attitudes toward immigration in an affluent, highly diverse, and multicultural society. Overall, a large majority of respondents acknowledge that immigration enriches Luxembourg's national identity, has a positive impact on the economy, and is not too large. Key findings include the central role of the respondent's foreign background in shaping such attitudes. We find that second-generation immigrants are more pro-immigration than those without immigrant background. Specifically, second-generation immigrants are significantly more positive about the cultural and economic contributions of immigration than are other natives. Pro-immigration attitudes are even more pronounced among first-generation immigrants. In particular, Portuguese immigrants, the largest country of origin, have the most positive attitudes towards the identity and economic effects of immigration, followed by non-EU immigrants and immigrants from bordering countries. Interestingly, with regard to the attitude towards the current level of immigration, we find a positive association with second-generation and first-generation background from bordering countries, but negative attitudes among non-EU immigrants. Immigrants from Portugal and other EU countries show no discernible differences to native-born on this question.

Our research also shows that exposure to immigration in the immediate neighborhood has little effect on the perceived cultural and economic benefits of immigration. However, local exposure to recent immigration flows influences attitudes about the optimal level of immigration among natives. A plausible rationale for these findings lies in the concept that embracing diversity and multiculturalism requires time for mutual adjustment by newcomers and their host community. We further provide suggestive evidence that natives' perceptions of the optimal level of immigration are mostly influenced by local exposure to non-European immigrants. However, the absence of long-lasting effects associated with the total share of immigrants suggests that ongoing interactions with non-EU immigrants may gradually mitigate the potential emergence of anti-migration sentiments over a decade or so.

Appendix

3.A Sample selection and representativeness

We recognize that our analysis is based on double selection: first, selection into responding to the survey; second, selection into consenting to linking the survey responses to the respondent's administrative data in the IGSS. This section discusses similarities and differences of the selected sample to the full survey sample and when the sample is weighted by population weights as discussed in Bienvenue et al. (2021).[40] Figure 3.A.1 shows the shares in each sample by background characteristics, in our analysis sample ("consent") as well as the weighted and unweighted shares in the full sample.

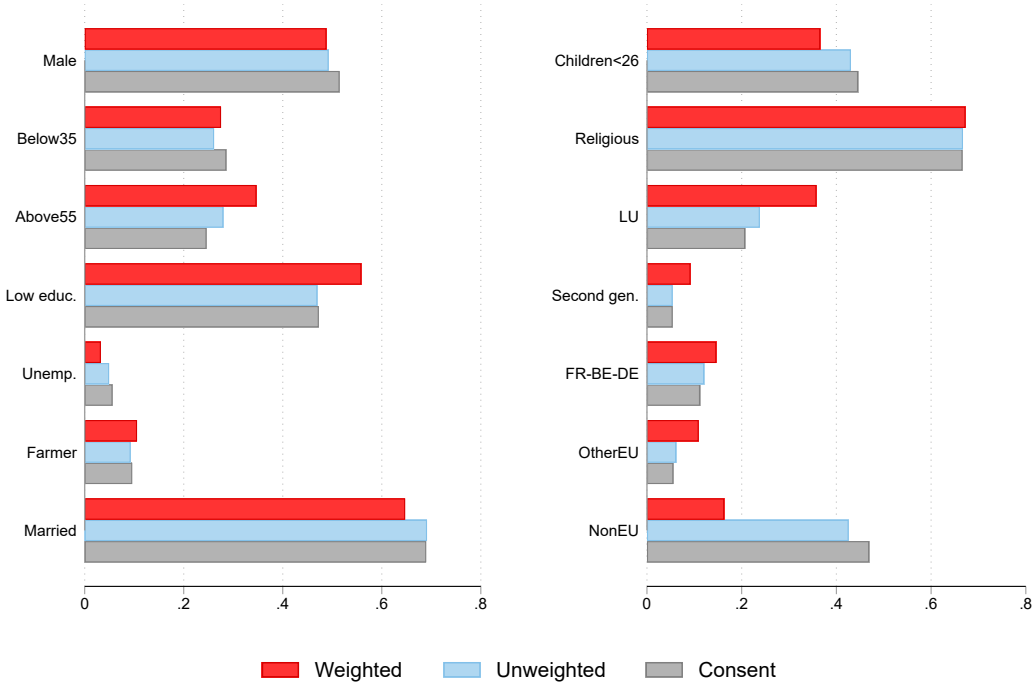


Figure 3.A.1: Differences in covariate distribution in weighted, unweighted and consent sample

Most strikingly, the graph shows that non-EU nationals were oversampled in the survey, while the Luxembourg-born without any migration background ("LU") are undersampled. We further observe higher shares of married respondents and those

[40] These population weights are only valid for the full sample, and are calculated by response rates and target rates in the population.

with children, while older individuals and those with lower educational background are undersampled. Some differences between the unweighted survey sample and the consent sample also emerge, albeit to a smaller extent.

To further investigate how selection into consent is correlated with these sociodemographic characteristics, Figure 3.A.2 provides estimates of the unconditional mean of the consent variable by each covariate. Respondents that consented to the use of their social security data are significantly more likely to be male, younger than 35, unemployed, have children, and be from a non-EU background. On the other hand, they are less likely to be born in Luxembourg to Luxembourgish parents.

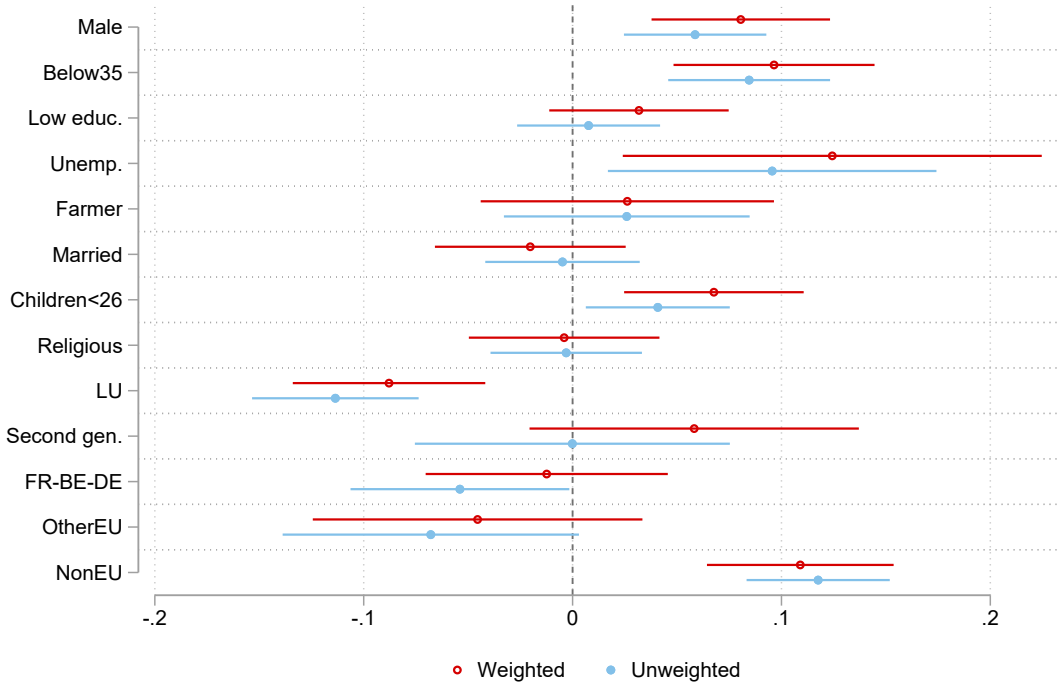


Figure 3.A.2: Selection into consent by covariates (unconditional mean)

Given the importance of migration (and native) background in our analysis, we further show how consent is correlated with the outcome variables of interest by origin. The coefficients in Figure 3.A.3 are based on linear regressions of the consent variable on each outcome and the full set of sociodemographic variables. The results show that native respondents and those from the bordering countries who believe that immigrants enrich the identity of the country are more likely to agree to using their social security data. On the other hand, Luxembourgish and non-EU citizens that believe that immigrants make the economy better are more likely to consent. Other attitudes are not strongly correlated with giving consent for any group.

While these results imply selection into consent by certain sociodemographic characteristics and some dimensions of attitudes, it is reassuring that this selection is not consistent among all groups and overall attitudes. Nevertheless, the empirical analysis should be interpreted cautiously with these limitations in mind.

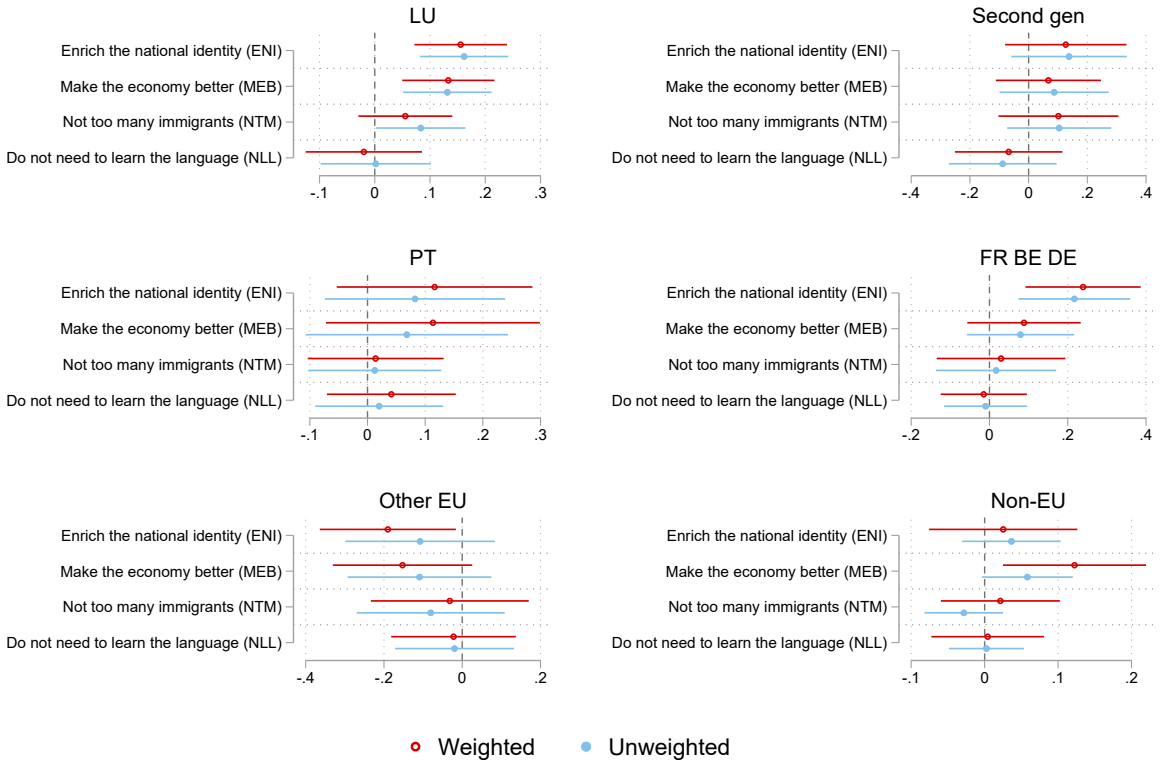


Figure 3.A.3: Selection into consent by attitude (conditional mean)

3.B Summary statistics

Table 3.B.1 provides descriptive statistics, distinguishing between the four dependent variables used in our empirical analysis (top panel), individual characteristics (middle panel), and neighborhood characteristics (bottom panel)

3.C Additional results

This section presents some supplementary results obtained when estimating Eq. (3.2).

3.C.1 Effect of migration background: Results for controls at the zip-code level

Table 3.4.1 showcases our regression of pro-immigration attitudes against individual attributes, while also accounting for neighborhood-level average characteristics calculated based at the zip code level. The influence of the latter is detailed in Table 3.C.1.

Table 3.B.1: Summary statistics

	Full sample	Born in Lux	Immigrant
Outcome Variables			
ENI	0.83	0.77	0.85
	(0.376)	(0.418)	(0.358)
MEB	0.81	0.76	0.83
	(0.393)	(0.426)	(0.379)
NTM	0.71	0.74	0.70
	(0.453)	(0.438)	(0.458)
NLL	0.35	0.18	0.42
	(0.478)	(0.381)	(0.493)
Individual Variables			
Male	0.52	0.52	0.52
	(0.500)	(0.500)	(0.500)
Below35	0.29	0.32	0.28
	(0.452)	(0.466)	(0.447)
Above55	0.24	0.35	0.20
	(0.429)	(0.479)	(0.403)
LS	0.08	0.07	0.08
	(0.270)	(0.261)	(0.273)
HS	0.61	0.48	0.66
	(0.488)	(0.500)	(0.475)
Unemployed	0.06	0.02	0.07
	(0.231)	(0.124)	(0.256)
Farmer or blue collar	0.10	0.05	0.11
	(0.295)	(0.228)	(0.314)
Married	0.61	0.50	0.65
	(0.488)	(0.500)	(0.477)
Children Below 26	0.45	0.32	0.49
	(0.497)	(0.468)	(0.500)
Religious	0.67	0.68	0.66
	(0.472)	(0.468)	(0.473)
Second gen immigrant	0.05	0.21	
	(0.227)	(0.406)	
FR-BE-DE	0.11		0.15
	(0.316)		(0.359)
PT	0.10		0.14
	(0.302)		(0.344)
OtherEU	0.06		0.08
	(0.230)		(0.264)
NonEU	0.47		0.64
	(0.499)		(0.481)
Neighborhood Variables			
Imm share zip	0.58	0.48	0.62
	(0.195)	(0.174)	(0.187)
Log Median Income	8.40	8.44	8.38
	(0.325)	(0.298)	(0.332)
Log Revis Beneficiaries	1.54	1.31	1.62
	(1.183)	(1.090)	(1.205)
Log Population	5.34	5.20	5.38
	(0.810)	(0.803)	(0.807)
Observations	1954	511	1443

Table 3.C.1: Effect of zip-code level characteristics on pro-immigration attitudes (OLS)

	ENI	MEB	NTM	NLL
Imm share zip	-0.007	-0.014	0.167	-0.032
	(0.084)	(0.090)	(0.102)	(0.106)
Δ Imm	-0.005	-0.004	-0.007*	-0.002
	(0.003)	(0.004)	(0.004)	(0.004)
Log Median Income	-0.019	-0.003	-0.078	-0.007
	(0.047)	(0.050)	(0.057)	(0.059)
Log Revis Beneficiaries	-0.022*	-0.010	-0.003	0.002
	(0.013)	(0.014)	(0.015)	(0.016)
Log Population	0.021	0.021	0.024	-0.026
	(0.017)	(0.018)	(0.020)	(0.021)
Observations	1876	1876	1876	1876
Individual controls	✓	✓	✓	✓

Notes: *** p<0.01, ** p<0.05, * p<0.10. The regressions show the estimates of the effect of zip-code level characteristics on opinions toward immigration as captured by the dependent variables EI, MEB, NTM, NLL. The dependent variables are defined as dummies that capture positive, more tolerant and less assimilationist opinions towards immigration. The regressions control for a set of individual-level characteristics (gender, age, education level, being married, having children, blue collar or farmer, being religious, being unemployed, migration background and origin dummies). All regressions are based on the full sample (Luxembourg- and foreign-born) and include a locality fixed effect. Standard errors are clustered at the locality level and bootstrapped when focusing on the number of recent immigrants.

3.C.2 Effects of individual level characteristics on attitudes towards immigration

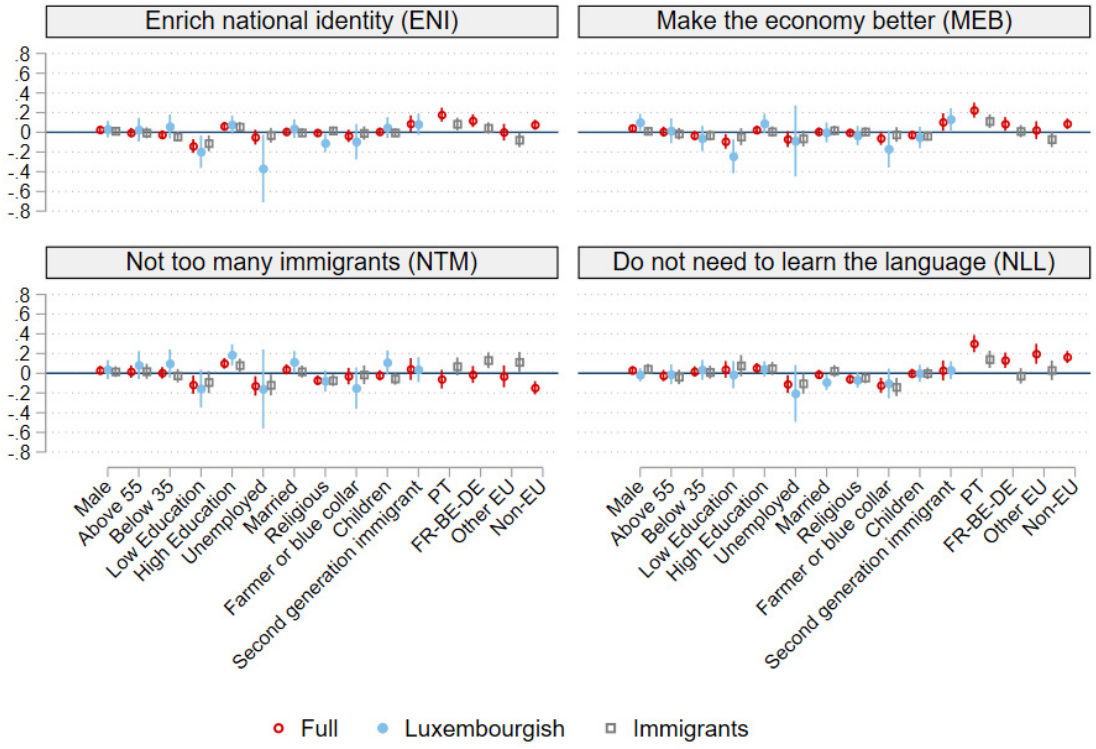
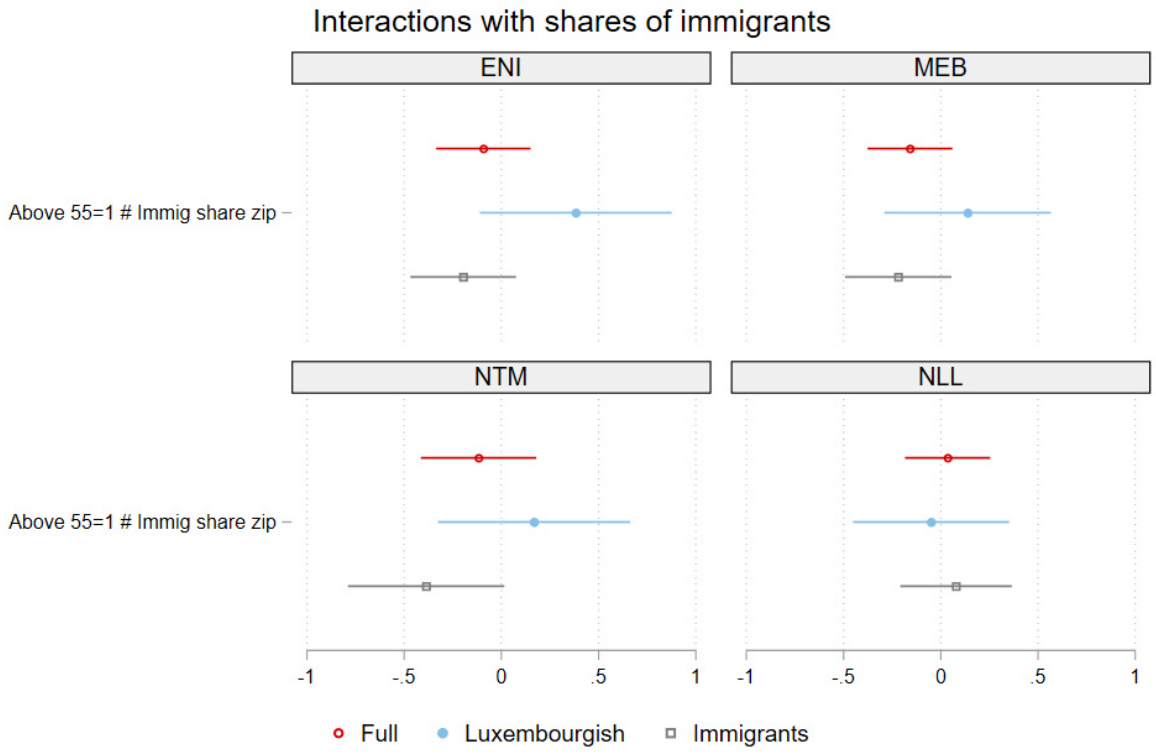


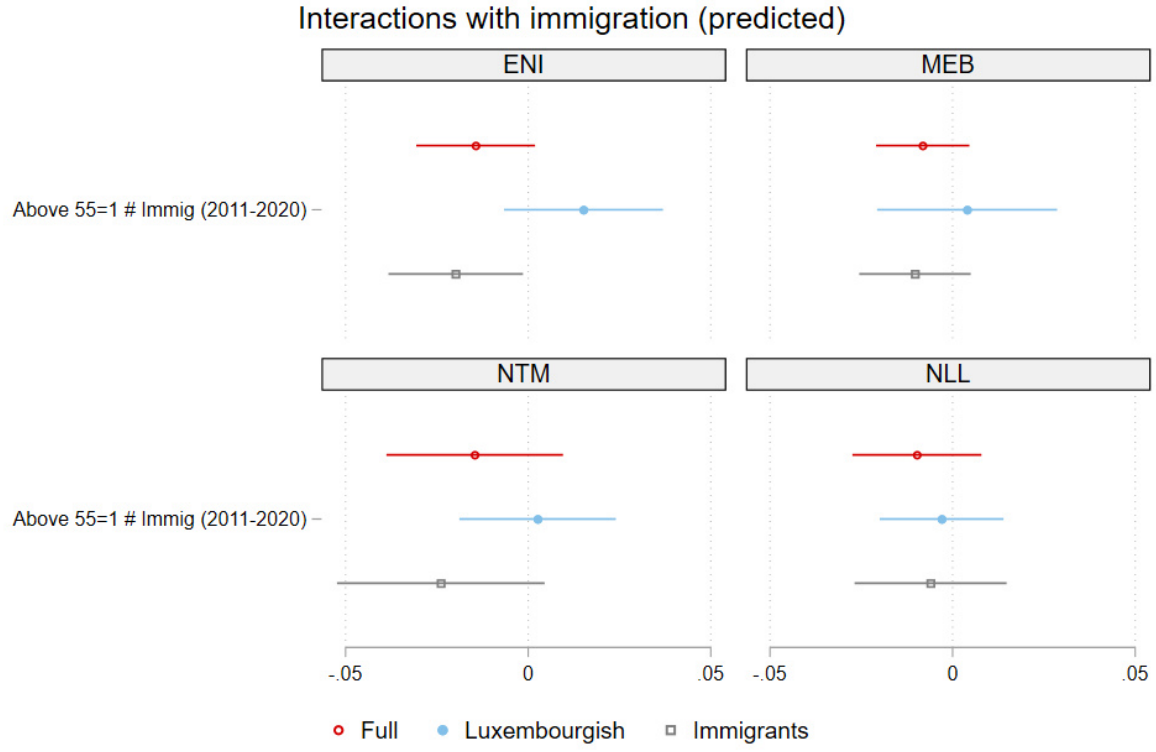
Figure 3.C.1: OLS effects of individual level characteristics on attitudes

3.C.3 Heterogeneity Analysis: Interaction with individual level characteristics

In this section, we add an interaction term between the share of immigrants and the recent immigration at the zip code level and some individual characteristics capturing old age, education, inactivity, and immigration background. We do not find evidence that any of this groups drive the effects.

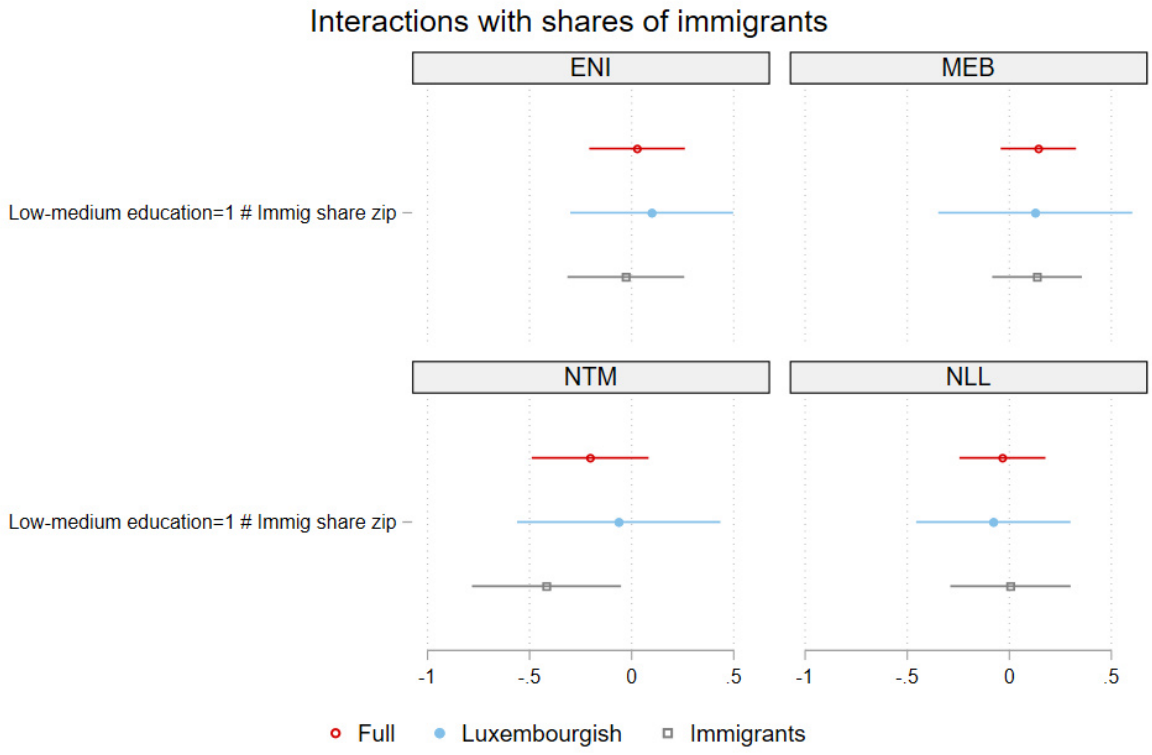


(a) Effect of immigrant share: Heterogeneity analysis with age (Above 55 Years Old).

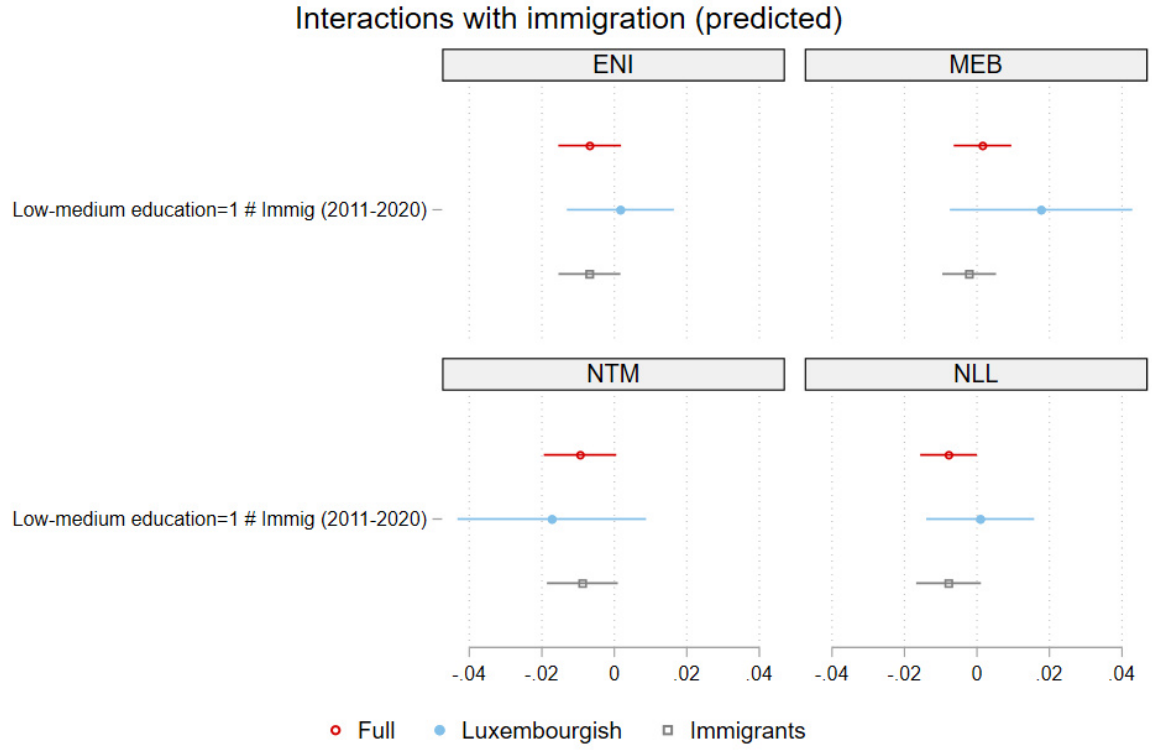


(b) Effect of recent immigrant waves: Heterogeneity analysis with age (Above 55 Years Old).

Figure 3.C.2: Heterogeneity analysis: Interaction with age



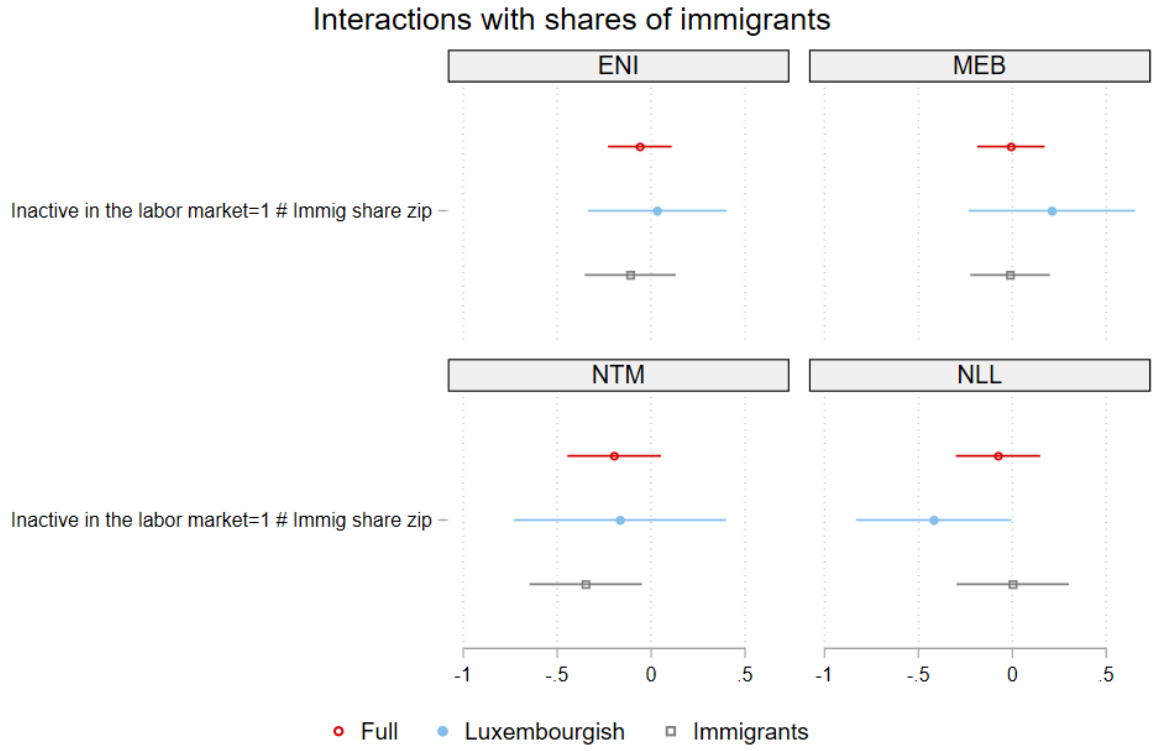
(a) Effect of immigrant share: Heterogeneity analysis with low education



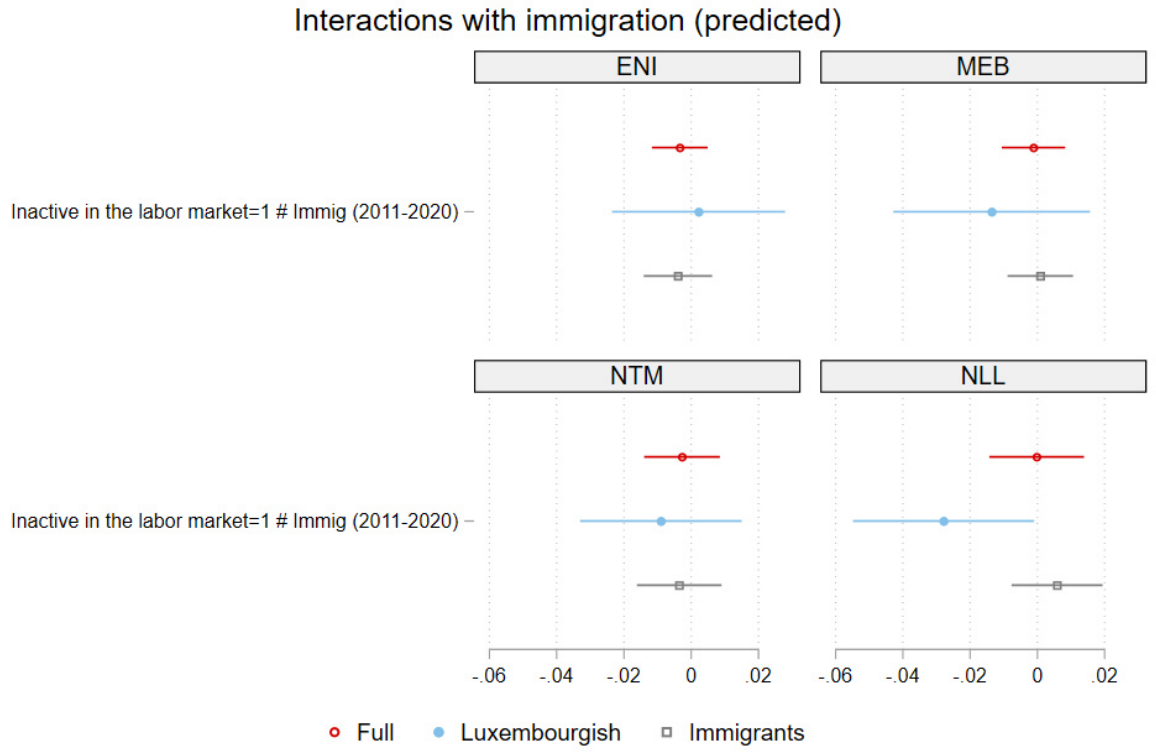
(b) Effect of recent immigrant waves: Heterogeneity analysis with low education

Figure 3.C.3: Heterogeneity Analysis: Interaction with education

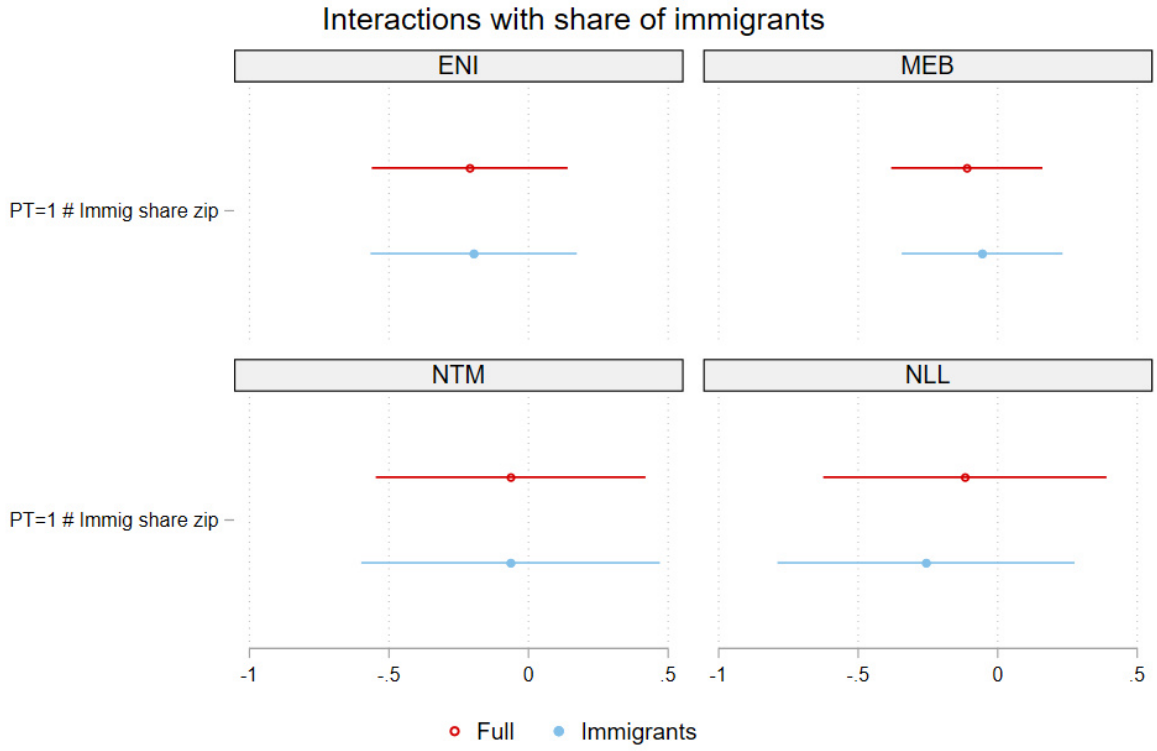
Figure 3.C.4: Heterogeneity analysis: Interaction with inactivity in the labor market



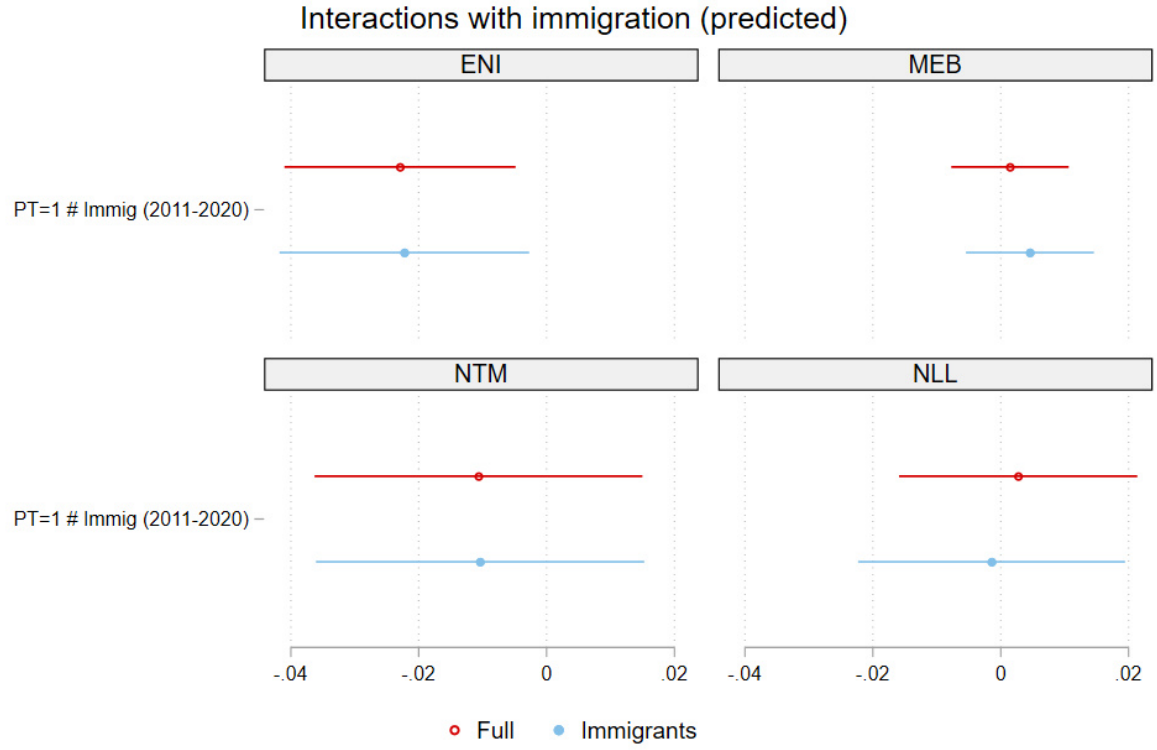
(a) Effect of immigrant share: Heterogeneity analysis with inactive status in the labor market



(b) Effect of recent Immigrant Waves: Heterogeneity analysis with with inactive status in the labor market

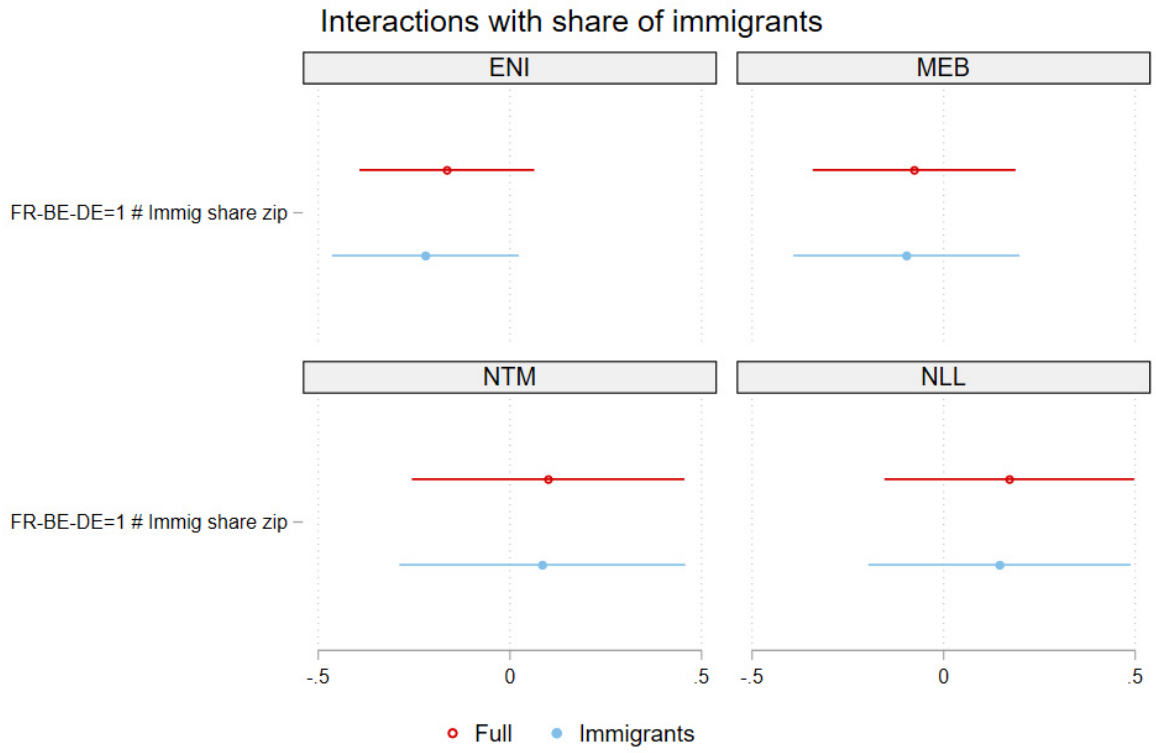


(c) Effect of immigrant share: Heterogeneity analysis with origin (Portuguese)

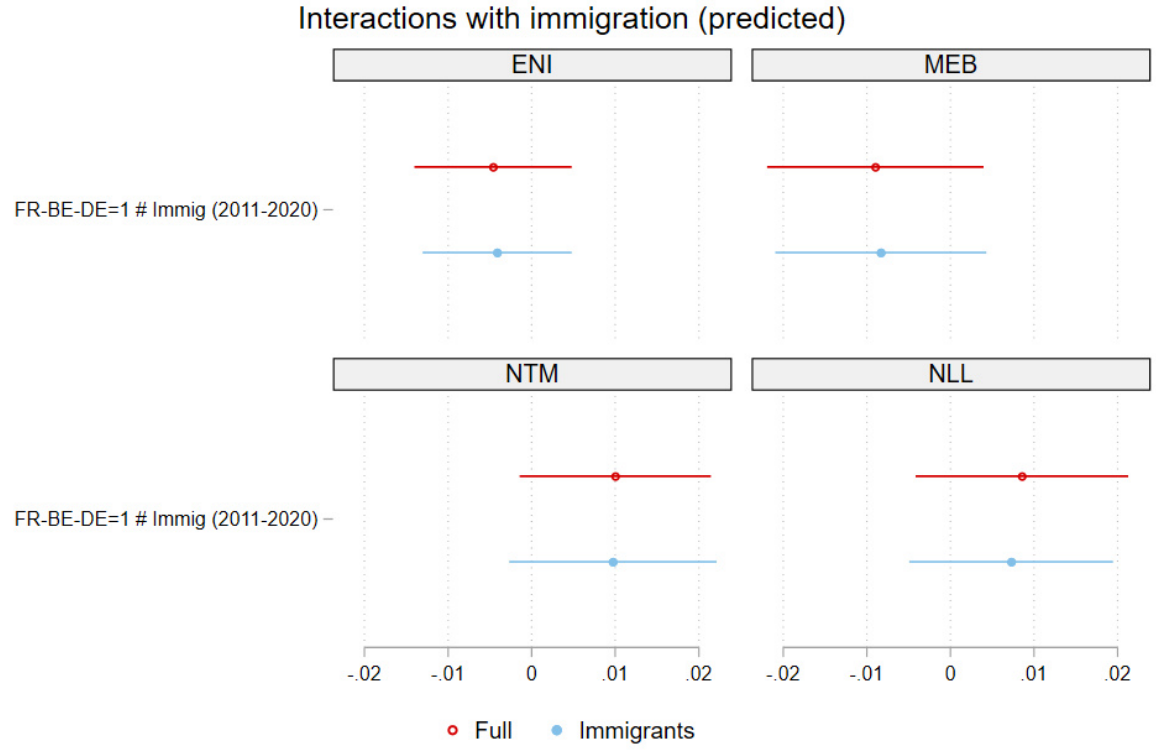


(d) Effect of recent immigrant waves: Heterogeneity analysis origin (Portuguese)

Figure 3.C.5: Heterogeneity analysis: Interaction with origin (Portuguese)

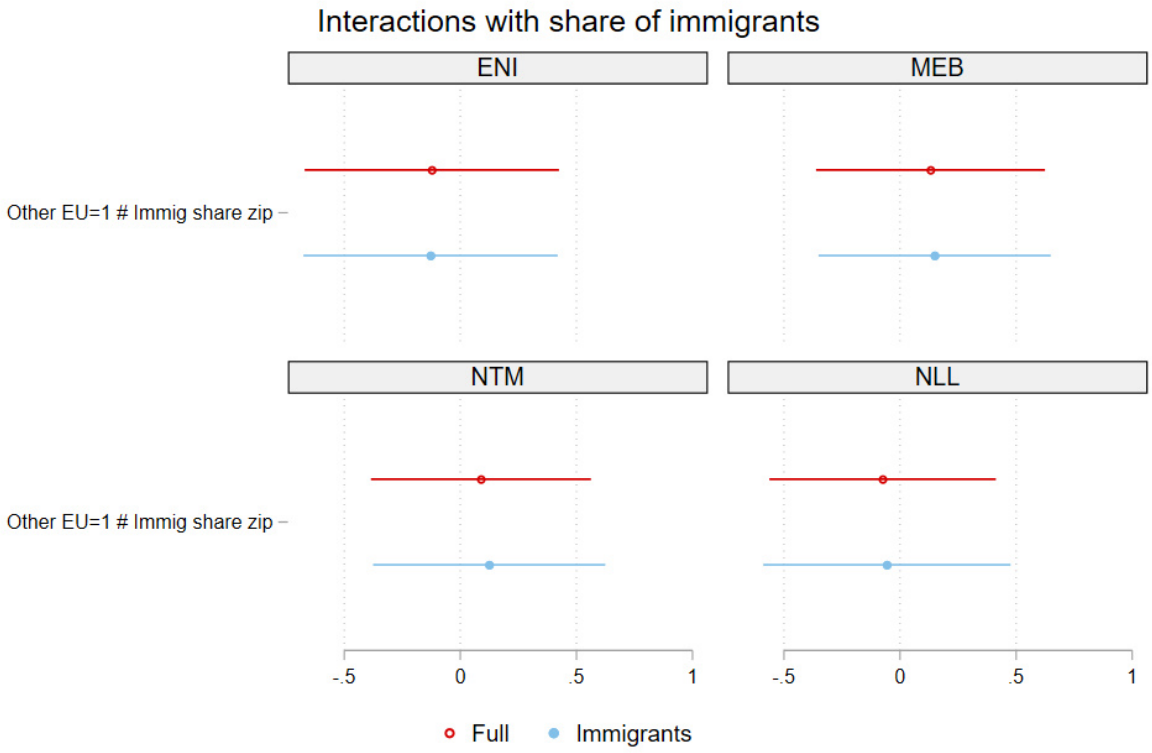


(a) Effect of immigrant share: Heterogeneity analysis with origin (FR-BE-DE)

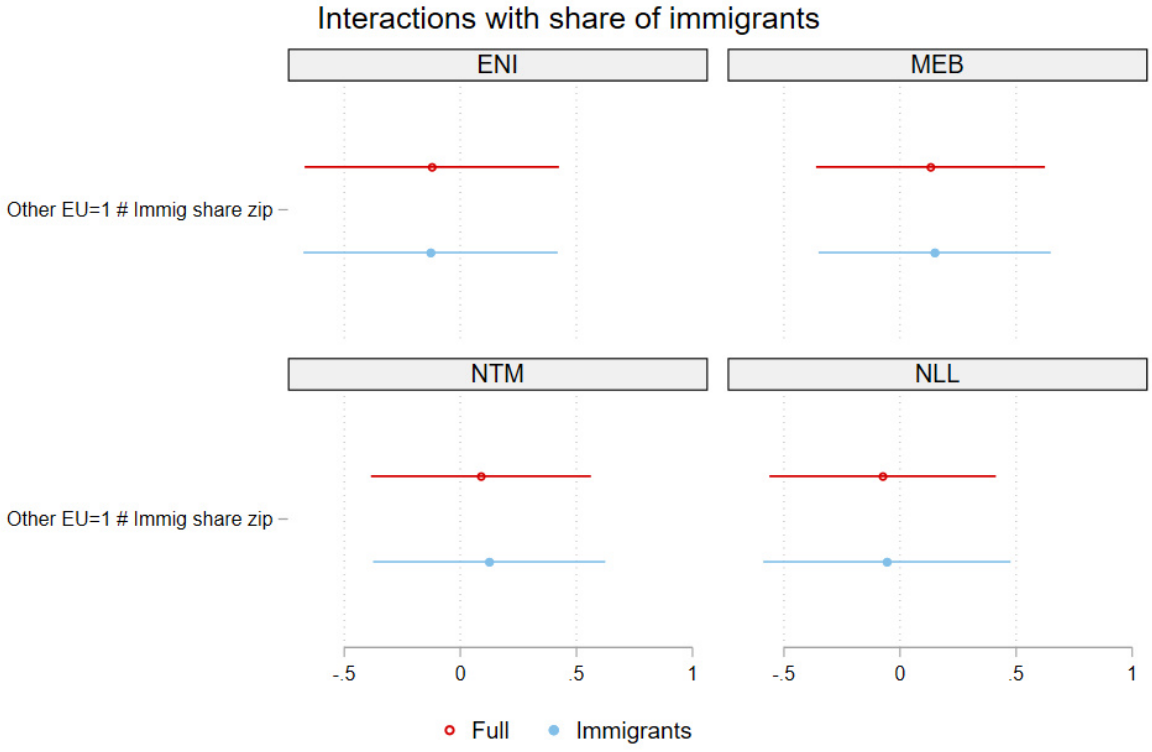


(b) Effect of recent immigrant waves: heterogeneity analysis origin (FR-BE-DE)

Figure 3.C.6: Heterogeneity analysis: Interaction with origin (FR-BE-DE)

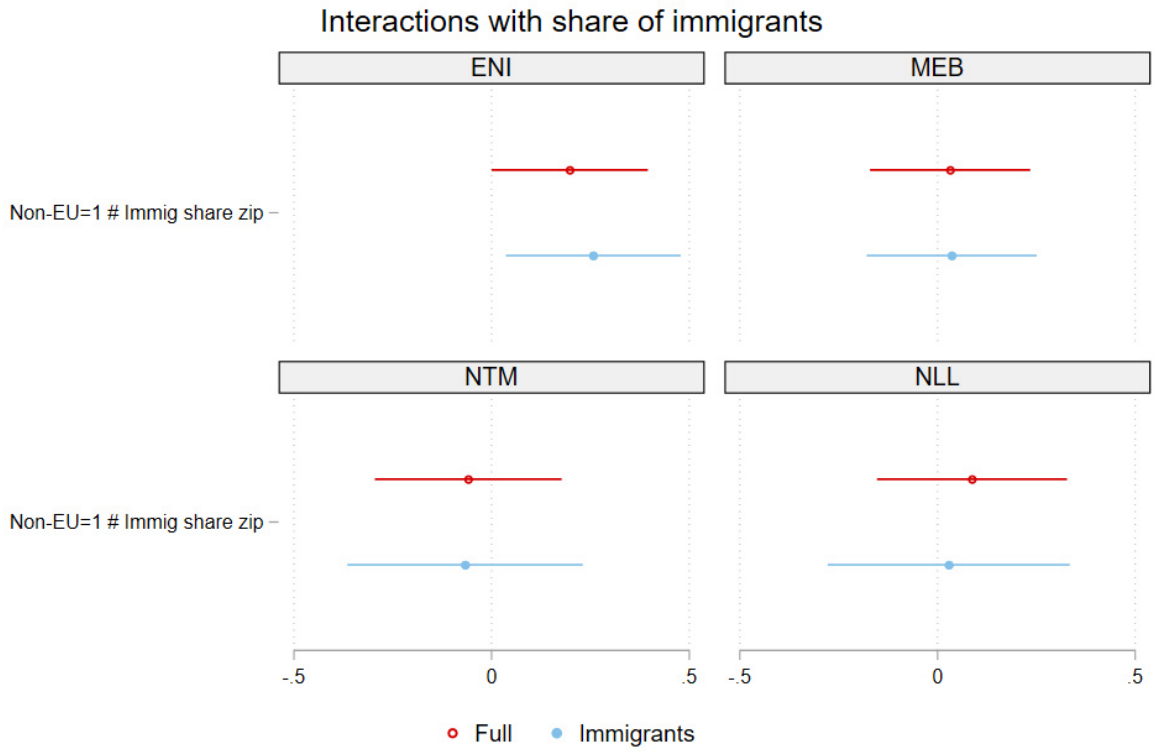


(a) Effect of immigrant share: Heterogeneity analysis with origin (Other EU)

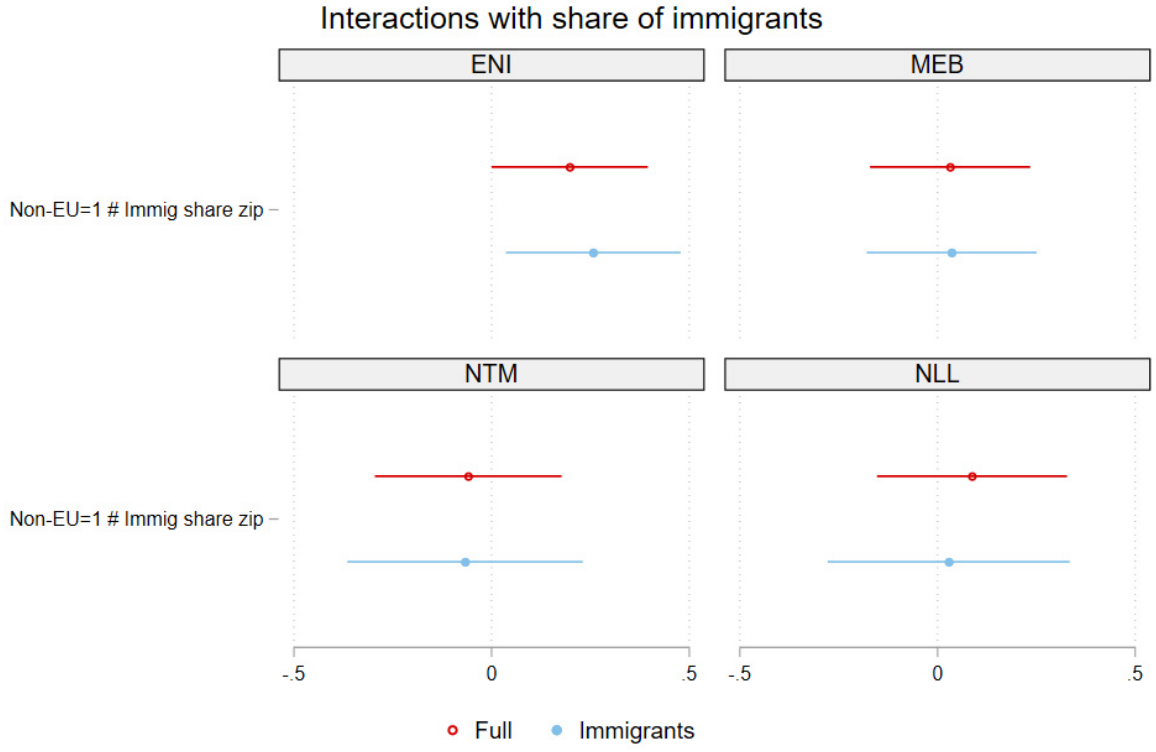


(b) Effect of recent immigrant waves: Heterogeneity analysis origin (Other EU)

Figure 3.C.7: Heterogeneity analysis: Interaction with origin (Other EU)

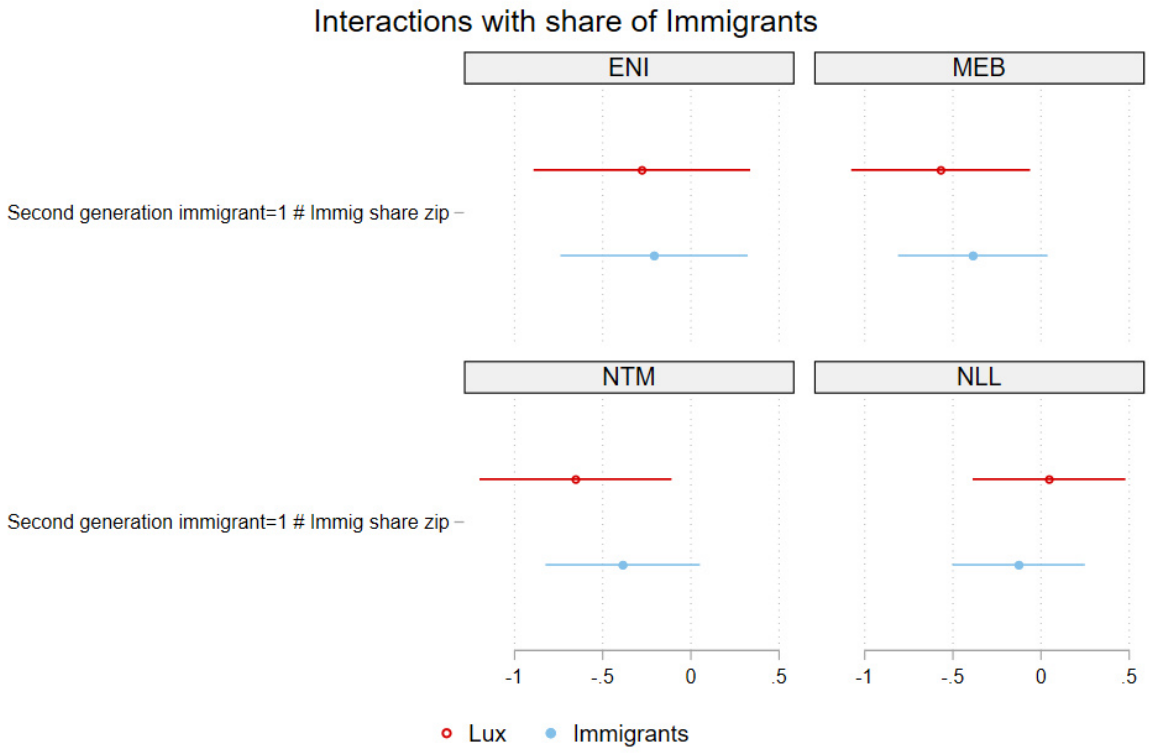


(a) Effect of immigrant share: Heterogeneity analysis with origin (Other EU)

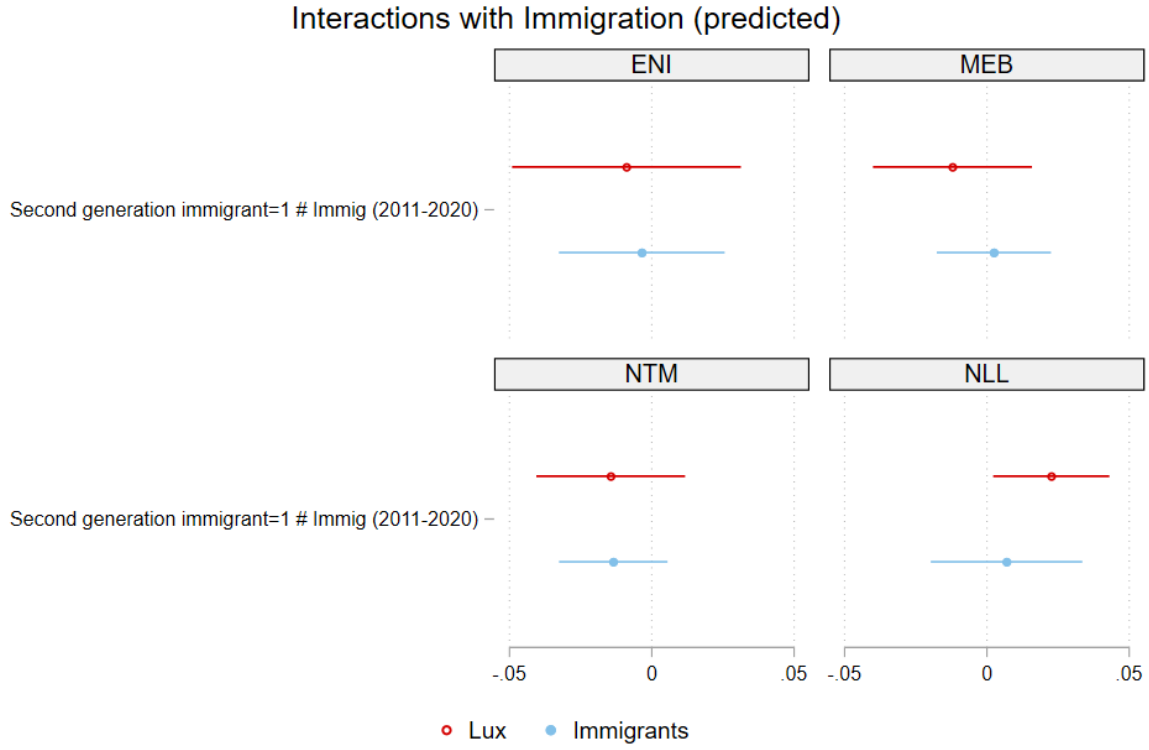


(b) Effect of recent immigrant eaves: Heterogeneity analysis origin (Other EU).

Figure 3.C.8: Heterogeneity analysis: Interaction with origin (Non-EU)



(a) Effect of immigrant share: Heterogeneity analysis with origin (Second generation immigrant)



(b) Effect of Recent immigrant waves: Heterogeneity analysis origin (Second generation immigrant).

Figure 3.C.9: Heterogeneity analysis: Interaction with origin (Second generation immigrant)

3.C.4 Bootstrapped standard errors

Table 3.C.2 provides estimates of the effect of changes in the number of immigrants at the zip code level on NTM, with alternative numbers of bootstrap replications.

Table 3.C.2: Effect on NTM with alternative numbers of bootstrap replications

	(1) No bootstrap	(2) 150 replications	(3) 300 replications	(4) 500 replications
Panel a				
Δ Immig (2011-2020)	-0.027***	-0.027**	-0.027**	-0.027**
	(0.010)	(0.012)	(0.012)	(0.012)
Observations	487	487	487	487
Panel b				
Δ Immig share (2011-2020)	-5.740***	-5.740**	-5.740**	-5.740**
	(2.094)	(2.313)	(2.385)	(2.486)
Observations	487	487	487	487
Controls	Yes	Yes	Yes	Yes
FE	Yes	Yes	Yes	Yes
Sample	Lux	Lux	Lux	Lux

Notes: *** p<0.01, ** p<0.05, * p<0.10. The regressions show the estimates of the effect the predicted change in immigration proxies on NTM. The dependent variable indicates tolerance with respect to the number of immigrants in the country. All the regressions restrict the sample to native respondents. Panel (a) shows the effect of the predicted change in the counts of immigrants by zip-code, between 2011 and 2020. Panel (b) shows the effect of the relative predicted change in immigration standardized by the total zip-code population. Col. (1) reports cluster robust standard errors. The other columns report bootstrapped standard errors. In Col. (2) the number of bootstrap replications is set to 150, in Col. (3) to 300 and in Col. (4) to 500. All regressions control for a set of individual-level characteristics (gender, age, education level, being married, having children, blue collar or farmer, being religious, being unemployed, migration background) as well as a set of zip-code characteristics (log median income, log share of individuals receiving social benefits, log of population).

3.D Pro-immigration index

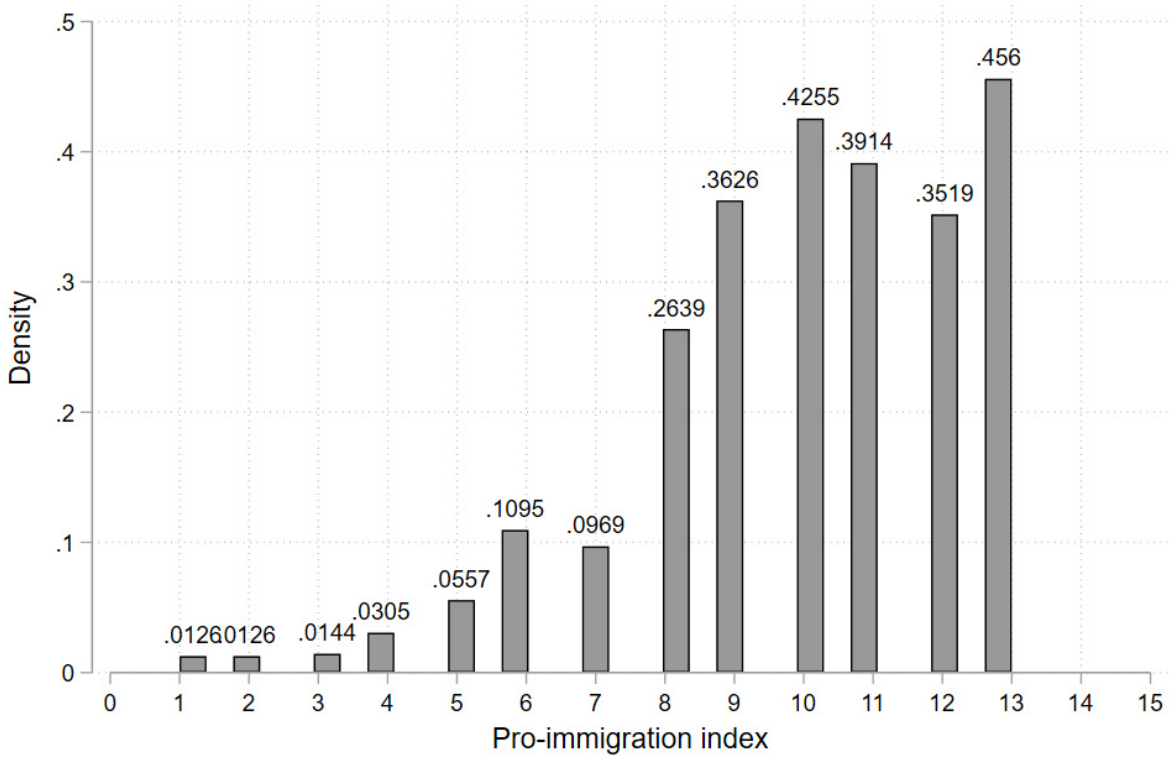
We also develop a pro-immigration index (labeled *Pro-immig*) that summarizes the immigration views expressed in the survey. To construct this index index, we use the Likert scaling method, creating a 4-point index for each question before aggregating them. For statements such as "Immigration enriches the country's identity," "Immigrants in Luxembourg improve life there," and "Luxembourg relies on immigrants to maintain its economy and social protection system," we assign a value of four for full agreement, three for partial agreement, two for partial disagreement, and one for full disagreement. Conversely, for the statement "There are too many immigrants in Luxembourg," we assign a value of one for full agreement and four for full disagreement. Our *Pro-immig* index is the sum of these four indices. Respondents who did not answer or said they did not know are treated as missing values. Stylized perceptions of the pro-immigration index are consistent with those of economic and cultural perceptions of immigration. In addition, we show that the empirical analysis of the determinants of the *Pro-immig* index mirrors that of the ENI and MEB variables.

3.E Exposure to immigration at the workplace

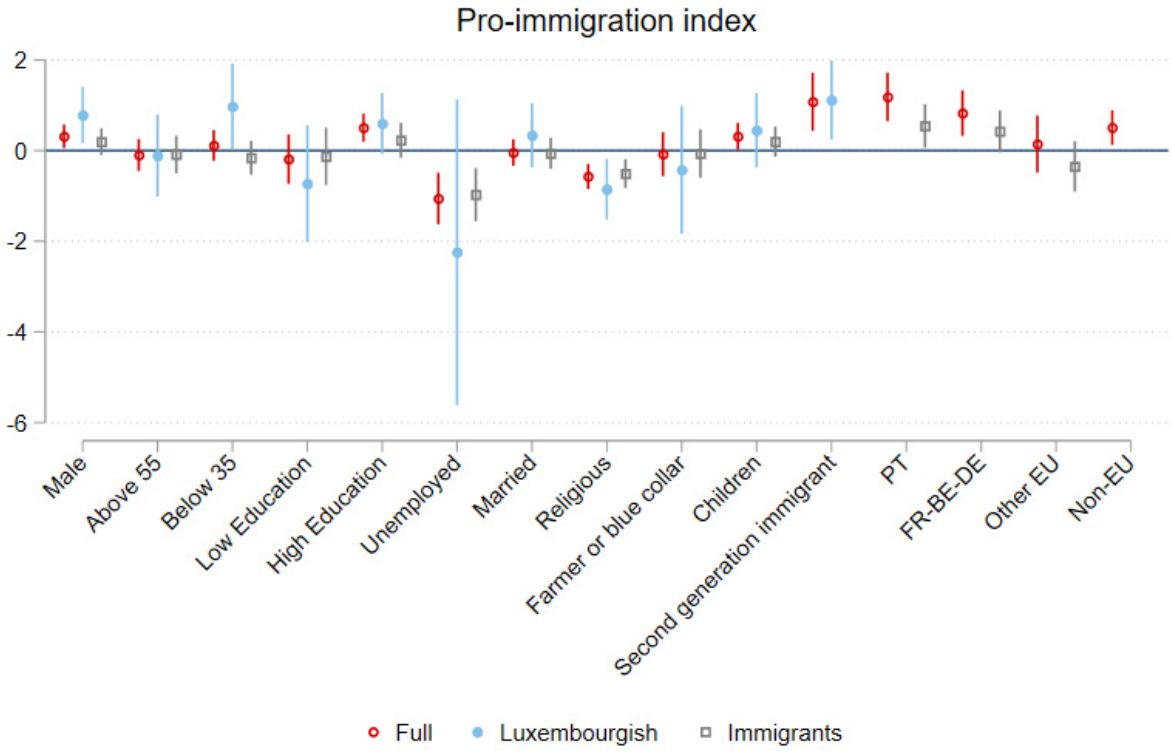
In this section, we adopt a comparable specification to that in Eq. (3.2), with a specific emphasis on the influence of the birthplace composition of the workforce within the main employing firm. Thus, the variable $\text{Immig}_{f,i}$ represents the proportion of immigrants in the total workforce of firm f of individual i as of 2020. For this regression analysis, we narrow our sample to respondents who are employed. This adjustment results in a sample size reduction by half compared to the one used in our baseline analysis (1,876 observations), yielding 947 observations. Given the challenge of identifying an appropriate instrument for this variable, we solely rely on OLS estimations.

Collectively, our findings do not unveil robust and substantial effects of workplace-based immigration exposure. Across the entire sample, the presence of immigrants at the firm level tends to elevate pro-immigration attitudes (ENI and NLL). When focusing on the native-born subset, a positive impact on MEB emerges, significant at the 5% level, and primarily driven by European migrants. Results are presented in Table 3.E.1.

While we have included a comprehensive set of individual and firm-level characteristics, as well as industry fixed effects, the lack of a more robust identification strategy requires a cautious interpretation of these estimates. Indeed, the estimates for firm-level exposure shares may be more consistent with the contact hypothesis than those for neighborhood exposure due to potential selection bias. Another interpretation to consider, however, is that workplaces may naturally satisfy Allport's conditions for effective intergroup contact. Specifically, these conditions include shared goals, intergroup cooperation, institutional support, and equal status among individuals, which are more likely to occur in the workplace than in the neighborhood.



(a) Pro-immigration index distribution



(b) Effect of Individual Level Characteristics on *Pro-Immig*

Figure 3.D.1: Pro-immigration index

Table 3.E.1: Effects of firm exposure to immigration

	(1) ENI	(2) MEB	(3) NTM	(4) NLL	(5) ENI	(6) MEB	(7) NTM	(8) NLL	(9) ENI	(10) MEB	(11) NTM	(12) NLL
Panel a												
Immig share firm	0.192***	0.108	0.070	0.219*	0.252	0.353*	-0.180	-0.259	0.145	0.016	0.183	0.315**
	(0.068)	(0.067)	(0.079)	(0.112)	(0.156)	(0.207)	(0.270)	(0.192)	(0.105)	(0.089)	(0.118)	(0.132)
Observations	947	947	947	947	222	222	222	222	725	725	725	725
Panel b												
EU share firm	0.218**	0.129*	0.062	0.140	0.229	0.405*	-0.190	-0.273	0.167	0.028	0.190	0.226*
	(0.083)	(0.077)	(0.086)	(0.109)	(0.183)	(0.227)	(0.308)	(0.189)	(0.121)	(0.097)	(0.115)	(0.130)
Non EU share firm	0.147*	0.072	0.083	0.357***	0.395	0.036	-0.121	-0.172	0.116	0.000	0.174	0.429***
	(0.080)	(0.075)	(0.106)	(0.130)	(0.353)	(0.339)	(0.358)	(0.349)	(0.109)	(0.094)	(0.142)	(0.141)
Observations	947	947	947	947	222	222	222	222	725	725	725	725
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE	Sector	Sector	Sector	Sector	Sector	Sector	Sector	Sector	Sector	Sector	Sector	Sector
Sample	Full	Full	Full	Full	Lux	Lux	Lux	Lux	Imm	Imm	Imm	Imm

Notes: *** p<0.01, ** p<0.05, * p<0.10. The regressions show the estimates of the effect exposure to immigration in the firm on opinions toward immigration as captured by the dependent variables EI, MEB, NTM, NLL. The dependent variables are defined as dummies that capture positive, more tolerant and less assimilationist opinions towards immigration. The first four columns use the full sample of respondents, the four columns in the middle restrict the sample to Luxembourgish respondents, the last four restrict the sample to foreign-born respondents. Immig share firm, EU share firm and Non-EU share firm are the proportions of immigrant EU and non-EU immigrants working in a firm, relative to the total workforce of the firm. The regressions control for a set of individual-level characteristics (gender, age, education level, being married, having children, blue collar or farmer, being religious, being unemployed, migration background and when the sample includes the foreign-born, origin dummies) as well as a set of firm-characteristics (firm size, average wage in the firm and share of blue collar workers). All regressions include a sector fixed effect. Standard errors are clustered at the sector level.

Bibliography

Adda, J., Pinotti, P., & Tura, G. (2025). There's More to Marriage Than Love: The Effect of Legal Status and Cultural Distance on Intermarriages and Separations [Publisher: University of Chicago Press]. *Journal of Political Economy*, *133*(4), 1276–1333.

Ahn, S. Y. (2021, September). *Matching Across Markets: An Economic Analysis of Cross-Border Marriage* (HCEO Working Paper Series No. 2021-047). The University of Chicago. Chicago IL.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, *8*(2), 155–194.

Alesina, A., Miano, A., & Stantcheva, S. (2023). Immigration and redistribution. *The Review of Economic Studies*, *90*(1), 1–39.

Alesina, A., Murard, E., & Rapoport, H. (2021). Immigration and preferences for redistribution in europe. *Journal of Economic Geography*, *21*(6), 925–954.

Alesina, A., & Tabellini, M. (2022). The political effects of immigration: Culture or economics? *NBER Working Paper*.

Algan, Y., Hémet, C., & Laitin, D. (2016). The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation. *Journal of Political Economy*, *124*(3), 696–733.

Allport, G. (1954). The nature of prejudice. *Cambridge: Cambridge University Press*.

Amirapu, A., Asadullah, M. N., & Wahhaj, Z. (2022). Social barriers to female migration: Theory and evidence from Bangladesh. *Journal of Development Economics*, *158*, 102891.

Anderberg, D. (2004). *Getting Hitched: The equilibrium Marriage Market Behaviour of a British Cohort* (Royal Holloway University of London Discussion Paper Series No. 2004-12). University of London.

Anderberg, D., Bagger, J., Bhaskar, V., & Wilson, T. (2019, March). *Marriage Market Equilibrium, Qualifications, and Ability* (IZA Discussion Paper No. No. 12210). Institute of Labor Economics.

Ashraf, N., Bau, N., Nunn, N., & Voena, A. (2020). Bride Price and Female Education. *Journal of Political Economy*, *128*(2), 591–641.

Bandyopadhyay, S., & Green, E. (2021). Explaining inter-ethnic marriage in sub-saharan africa. *Journal of International Development*, *33*(4), 627–643.

Bartik, T. (1991). Who benefits from state and local economic development policies? *Books from Upjohn Press, W.E. Upjohn Institute for Employment Research*.

Basu, S. (2015). INTERMARRIAGE AND THE LABOR MARKET OUTCOMES OF ASIAN WOMEN. *Economic Inquiry*, *53*(4), 1718–1734.

Bau, N. (2021). Can Policy Change Culture? Government Pension Plans and Traditional Kinship Practices. *American Economic Review*, *111*(6), 1880–1917.

Bauer, T., Lofstrom, M., & Zimmermann, K. (2000). Immigration policy, assimilation of immigrants, and natives' sentiments towards immigrants: Evidence from 12 oecd countries. *Swedish Economic Policy Review*, *7*, 11–53.

Bayer, P., Ross, S., & Topa, G. (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of political Economy*, *116*(6), 1150–1196.

Bazzi, S., Gaduh, A., Rothenberg, A., & Won, M. (2019). Unity in diversity? how intergroup contact can foster nation building. *American Economic Review*, *109*(11), 3978–4025.

Bazzi, S., Gaduh, A., Rothenberg, A. D., & Wong, M. (2016). Skill Transferability, Migration, and Development:Evidence from Population Resettlement in Indonesia. *American Economic Review*, *106*(9), 2658–2698.

Becerra-Valbuena, L. G., & Millock, K. (2021). Gendered migration responses to drought in Malawi. *Journal of Demographic Economics*, *87*(3), 437–477.

Becker, G. S. (1973). A theory of marriage: Part i. *Journal of Political economy*, *81*(4), 813–846.

Becker, G. S. (1974). A theory of marriage: Part ii. *Journal of Political Economy*, *82*(2), 11–26.

Belot, M., & Francesconi, M. (2013). Dating Preferences and Meeting Opportunities in Mate Choice Decisions [Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System]]. *The Journal of Human Resources*, *48*(2), 474–508.

Bertinelli, L., Cömertpay, R., & Maystadt, J.-F. (2025). Ethnic diversity and conflict in sub-Saharan Africa: Evidence from refugee-hosting areas. *Journal of Development Economics*, *172*, 103393.

Bienvenue, J., Docquier, F., Guadarrama Sanz, M., Tenikue, M., & Uljarevic, A. (2021). *Enquete sur le racisme et les discriminations ethno-raciales au luxembourg* (tech. rep.). Report: Ministry of Family, Integration and the Greater Region.

Bloom, D. E., & Khanna, T. (2007). The Urban Revolution. *Finance and Development*, *44*(3).

Blossfeld, H.-P. (2009). Educational Assortative Marriage in Comparative Perspective. *Annual Review of Sociology*, *35*(1), 513–530.

Blossfeld, H.-P., & Timm, A. (2003). Educational Systems as Marriage Markets in Modern Societies: A Conceptual Framework. In H.-P. Blossfeld & A. Timm (Eds.), *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies* (pp. 1–18). Springer Netherlands.

Boesen, E. (2020). Moving from nation into region. experiences and memories of cross-border dwelling in the greater region saarlorlux. In *Border experiences in europe. everyday life – working life – communication – languages*. Nomos.

Bryan, G., Chowdhury, S., & Mobarak, A. M. (2014). Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica*, *82*(5), 1671–1748.

Bryan, G., & Morten, M. (2019). The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia. *Journal of Political Economy*, *127*(5), 2229–2268.

Bursztyn, L., Chaney, T., Hassan, T. A., & Rao, A. (2024). The immigrant next door. *American Economic Review*, *114*(2), 348–384.

Card, D., Dustmann, C., & Preston, I. (2012). Immigration, wages and compositional amenities. *Journal of the European Economic Association*, *10*(1), 78–119.

Charles, K. K., Hurst, E., & Killewald, A. (2013). Marital Sorting and Parental Wealth [Publisher: Springer]. *Demography*, *50*(1), 51–70.

Choo, E., & Siow, A. (2006). Who marries whom and why. *Journal of political Economy*, *114*(1), 175–201.

Cinelli, C., Forney, A., & Pearl, J. (2024). A Crash Course in Good and Bad Controls. *Sociological Methods & Research*, *53*(3), 1071–1104.

Ciscato, E. (2024). Assessing Racial and Educational Segmentation in Large Marriage Markets. *Review of Economic Studies*, rdae115.

Cools, S., Finseraas, H., & Rogeberg, O. (2021). Local immigration and support for anti-immigration parties: A meta-analysis. *American Journal of Political Science*, *65*(4), 988–1006.

Corno, L., La Ferrara, E., & Burns, J. (2019). Interaction, stereotypes and performance. evidence from south africa. *IFS Working Paper*, *W19/03*.

Crespin-Boucaud, J. (2020). Interethnic and interfaith marriages in sub-Saharan Africa. *World Development*, *125*, 104668.

Debray, A., Ruyssen, I., & Schewel, K. (2025). The Aspiration to Stay: A Global Analysis. *International Migration Review*, *59*(3), 1385–1431.

Demarest, L., & Haer, R. (2022). A perfect match? The dampening effect of interethnic marriage on armed conflict in Africa. *Conflict Management and Peace Science*, *39*(6), 686–705.

Deming, W. E., & Stephan, F. F. (1940). On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. *The Annals of Mathematical Statistics*, *11*(4), 427–444.

Docquier, F., Guichard, L., Iandolo, S., Rapoport, H., Turati, R., & Vannoorenberghe, G. (2024). Populism and the skill-content of globalization: Evidence from the last 60 years. *CEPR Discussion Paper*, *18822*.

Docquier, F., Szymanska, A., Gerber, P., Ferro, Y., Schiel, K., & Pigeron-Piroth, I. (2023). Geographical distribution of immigrants in luxembourg: Dynamics and spatial segregation with natives. *STATEC RP 1st Results*, *9*.

Domènech-Arumí, G. (2025). Neighborhoods, perceived inequality, and preferences for redistribution: Evidence from barcelona. *Journal of Public Economics*, *242*, 105288.

Drazanova, L., Gonnot, J., Heidland, T., & Kruger, F. (2022). Understanding differences in attitudes to immigration: A meta-analysis of individual-level factors. *Manuscript*.

Dupuy, A. (2021). Migration in China: To work or to wed? *Journal of Applied Econometrics*, *36*(4), 393–415.

Dupuy, A., & Weber, S. (2022). Marriage Market Counterfactuals Using Matching Models. *Economica*, *89*(353), 29–43.

Easterly, W., & Levine, R. (1997). Africa's Growth Tragedy: Policies and Ethnic Divisions. *The Quarterly Journal of Economics*, *112*(4), 1203–1250.

Edo, A., Giesing, Y., Öztunc, J., & Poutvaara, P. (2019). Immigration and electoral support for the far left and far right. *European Economic Review*, *115*, 99–143.

Eika, L., Mogstad, M., & Zafar, B. (2019). Educational Assortative Mating and Household Income Inequality. *Journal of Political Economy*, *127*(6), 2795–2835.

Esteve, A., Schwartz, C. R., Van Bavel, J., Permanyer, I., Klesment, M., & García-Román, J. (2016). The End of Hypergamy: Global Trends and Implications. *Population and Development Review*, *42*(4), 615–625.

Facchini, G., & Mayda, A. (2017). Does the welfare state affect individual attitudes toward immigrants? evidence across countries. *European Journal of Political Research*, *56*(3), 703–722.

Farahzadi, S. (2024). Migration for Marriage. *Manuscript*.

Fielding, D. (2024). Community-level ethnic diversity and community-level socio-economic development: Evidence from 20 African countries. *New Zealand Economic Papers*, *58*(1), 1–19.

Florczyk, A. J., Corbane, C., Ehrlich, D., Freire, S., Kemper, T., Maffenini, L., Melchiorri, M., Micheleand Pesaresi, Politis, P., Schiavina, M., Sabo, F., & Zanchetta, L. (2019). *GHSL data package 2019: Public release GHS P2019*. Publications Office.

Galichon, A., & Salanié, B. (2022). Cupid's Invisible Hand: Social Surplus and Identification in Matching Models. *The Review of Economic Studies*, *89*(5), 2600–2629.

Galichon, A., & Salanié, B. (2024). Estimating separable matching models. *Journal of Applied Econometrics*, *39*(6), 1021–1044.

Gihleb, R., Giuntella, O., & Stella, L. (2022). Exposure to past immigration waves and attitudes toward newcomers. *IZA Discussion Paper Series*, *15545*.

Goldman, B., Gracie, J., & Porter, S. (2025). Who Marries Whom? The Role of Segregation by Race and Class.

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, *110*(8), 2586–2624.

Gollin, D., Kirchberger, M., & Lagakos, D. (2021). Do urban wage premia reflect lower amenities? Evidence from Africa. *Journal of Urban Economics*, *121*, 103301.

Gollin, D., Lagakos, D., & Waugh, M. E. (2014). The Agricultural Productivity Gap. *The Quarterly Journal of Economics*, *129*(2), 939–993.
Goux, D., & Maurin, E. (2007). Close neighbours matter: Neighbourhood effects on early performance at school. *Economic Journal*, *117*(523), 1193–1215.
Gray, C., & Mueller, V. (2012). Drought and Population Mobility in Rural Ethiopia. *World Development*, *40*(1), 134–145.
Guriev, S., & Papaioannou, E. (2022). The political economy of populism. *Journal of Economic Literature*, *60*(3), 753–832.
Hainmueller, J., & Hiscox, M. (2010). Attitudes toward highly skilled and low-skilled immigration: Evidence from a survey experiment. *American Political Science Review*, *104*(1), 61–84.
Hainmueller, J., & Hopkins, D. (2014). Public attitudes towards immigration. *Annual Reviews of Political Science*, *17*, 225–249.
Hamilton, G., & Siow, A. (2007). Class, gender and marriage. *Review of Economic Dynamics*, *10*(4), 549–575.
Hamon, R., & Ingoldsby, B. (2003). *Mate Selection Across Cultures*. SAGE Publications, Inc.
Hidrobo, M., Mueller, V., & Roy, S. (2022). Cash transfers, migration, and gender norms. *American Journal of Agricultural Economics*, *104*(2), 550–568.
Imbert, C., Monras, J., Seror, M., & Zylberberg, Y. (2023, December). *Floating population: Migration with(out) family and the spatial distribution of economic activity* (CEPR Discussion Paper No. No. 18418). CEPR Press. Paris & London.
Imbert, C., & Papp, J. (2020). Costs and benefits of rural-urban migration: Evidence from India. *Journal of Development Economics*, *146*, 102473.
Jaffe, S., & Weber, S. (2019). The effect of meeting rates on matching outcomes [Publisher: Springer]. *Economic Theory*, *67*(2), 363–378.
Kalmijn, M. (1998). Intermarriage and Homogamy: Causes, Patterns, Trends. *Annual Review of Sociology*, *24*, 395–421.
Kalmijn, M., & Flap, H. (2001). Assortative Meeting and Mating: Unintended Consequences of Organized Settings for Partner Choices [Publisher: Oxford University Press]. *Social Forces*, *79*(4), 1289–1312.
Katz, L., Kling, J., & Liebman, J. (2001). Moving to opportunity in boston: Early results of a randomized mobility experiment. *Quarterly Journal of Economic*, *116*, 607–654.
Kawaguchi, D., & Lee, S. (2017). Brides for Sale: Cross-Border Marriages and Female Immigration [Publisher: Wiley]. *Economic Inquiry*, *55*(2), 633–654.
Kling, J., Liebman, J., & Katz, L. (2007). Experimental analysis of neighborhood effects. *Econometrica*, *75*(1), 83–119.
Lagakos, D. (2020). Urban-Rural Gaps in the Developing World: Does Internal Migration Offer Opportunities? *Journal of Economic Perspectives*, *34*(3), 174–192.
Lagakos, D., Marshall, S., Mobarak, A. M., Vernot, C., & Waugh, M. E. (2020). Migration costs and observational returns to migration in the developing world. *Journal of Monetary Economics*, *113*, 138–154.
Lagakos, D., Mobarak, A. M., & Waugh, M. E. (2023). The Welfare Effects of Encouraging Rural–Urban Migration. *Econometrica*, *91*(3), 803–837.
Lam, D. (1988). Marriage Markets and Assortative Mating with Household Public Goods: Theoretical Results and Empirical Implications [Publisher: JSTOR]. *The Journal of Human Resources*, *23*(4), 462.
Levi, E., Mariani, R., & Patriarca, F. (2020). Hate at first sight? dynamic aspects of the electoral impact of migration: The case of ukip. *Journal of Population Economics*, *33*, 1–32.
Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review*, *111*(6), 1807–1844.
Luke, N., & Munshi, K. (2006). New roles for marriage in urban africa: Kinship networks and the labor market in kenya. *Review of Economics and Statistics*, *88*(2), 264–282.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, *60*, 531–542.
Mayda, A. (2006). Who is against immigration? a cross-country investigation of individual attitudes toward immigrants. *Review of Economics and Statistics*, *88*(3), 510–530.
Migration Policy Institute. (2024). International migrants by country of destination, 1960-2024 [https://www.migrationpolicy.org/programs/data-hub/charts/international-migrants-country-destination-1960-2024 [Accessed: 2025-10-02]].
Moriconi, S., Peri, G., & Turati, R. (2019). Immigration and voting for redistribution: Evidence from european elections. *Labour Economics*, *61*(101765).
Moriconi, S., Peri, G., & Turati, R. (2022). Skill of the immigrants and vote of the natives: Immigration and nationalism in european elections 2007–2016. *European Economic Review*, *141*, 103986.
Müller-Crepon, C., Pengl, Y. I., & Bormann, N.-C. (2021). Linking Ethnic Data from Africa (LEDA). *Journal of Peace Research*, *Online First*.
Murdock, G. P. (1959). *Africa. its peoples and their culture history*. McGraw-Hill Book Company.
NISO CRediT Working Group. (2022). ANSI/NISO Z39.104-2022, CRediT, Contributor Roles Taxonomy.
Nunn, N., & Wantchekon, L. (2011). The Slave Trade and the Origins of Mistrust in Africa. *American Economic Review*, *101*(7), 3221–3252.
Oreopoulos, P. (2003). The long-run consequences of living in a poor neighborhood. *Quarterly Journal of Economic*, *118*, 1533–1575.
Paluck, E., Green, S., & Green, D. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy*, *3*(2), 129–158.
Pengl, Y. I., Roessler, P., & Rueda, V. (2022). Cash Crops, Print Technologies, and the Politicization of Ethnicity in Africa. *American Political Science Review*, *116*(1), 181–199.
Pesando, L. M. (2021). Educational Assortative Mating in Sub-Saharan Africa: Compositional Changes and Implications for Household Wealth Inequality. *Demography*, *58*(2), 571–602.
Pettigrew, T., & Tropp, L. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783.
Robinson, A. L. (2020). Ethnic Diversity, Segregation and Ethnocentric Trust in Africa. *British Journal of Political Science*, *50*(1), 217–239.
Rosenzweig, M. R., & Stark, O. (1989). Consumption Smoothing, Migration, and Marriage: Evidence from Rural India. *Journal of Political Economy*, *97*(4), 905–926.
Ruggles, S., Cleveland, L. L., Lovatón Dávila, R., Sarkar, S., Sobek, M., Burk, D., Ehrlich, D. E., Heimann, Q., & Lee, J. (2024). Integrated Public Use Microdata Series, International: Version 7.5.
Sarracino, F., & Riillo, C. (2022). Facing the challenge of globalization: The role of confidence in institutions. *Scienza e Pace*, *13*(1), 85–136.
Schaner, S., & Das, S. (2016). *Female Labor Force Participation in Asia: Indonesia Country Study*. Asian Development Bank.
Schaub, M., Gereke, J., & Baldassarri, D. (2021). Strangers in hostile lands: Exposure to refugees and right-wing support in germany's eastern regions. *Comparative Political Studies*, *54*(3-4), 686–717.
Schneider-Strawczynski, S. (2021). *When is contact effective? evidence on refugee-hosting and far-right support in france* (tech. rep.). Manuscript.
Shapley, L. S., & Shubik, M. (1971). The assignment game I: The core. *International Journal of Game Theory*, *1*(1), 111–130.
Steinmayr, A. (2021). Contact versus exposure: Refugee presence and voting for the far right. *Review of Economics and Statistics*, *103*(2), 310–327.
Strauss, J., Witoelar, F., & Sikoki, B. (2016, March). The Fifth Wave of the Indonesia Family Life Survey (IFLS5): Overview and Field Report.

UNDESA. (2019, August). *World Urbanization Prospects: The 2018 Revision*. UN.
UNDP. (2009). *Human Development Report 2009: Overcoming Barriers - Human Mobility and Development*. United Nations Development Program.
Verheyden, B., Pigeron-Piroth, I., Dautel, V., Fehlen, F., Ferro, Y., & Schiel, K. (2024). Active residents: A dynamic and segmented labor market. *STATEC RP 1st Results*, *forthcoming*.
Wajdi, N., Mulder, C. H., & Adioetomo, S. M. (2017). Inter-regional migration in Indonesia: A micro approach. *Journal of Population Research*, *34*(3), 253–277.
Weiss, Y., Yi, J., & Zhang, J. (2018). CROSS-BORDER MARRIAGE COSTS AND MARRIAGE BEHAVIOR: THEORY AND EVIDENCE [Publisher: Wiley]. *International Economic Review*, *59*(2), 757–784.
Wen, J., & Mao, J. (2025). Assortative Marriage and Geographic Sorting *.
Wessel, T. (2009). Does diversity in urban space enhance intergroup contact and tolerance? *Geografiska Annaler: Series B, Human Geography*, *91*(1), 5–17.
Wong, L. Y. (2003). Structural Estimation of Marriage Models. *Journal of Labor Economics*, *21*(3), 699–727.
Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. The MIT Press.
Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach*. South-Western Pub, Mason.
Wooldridge, J. M., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.