

The Fabric of Movement: Linking Patterns, Drivers, and Mechanisms in International Migration

Micol Matilde Morellini

Nuffield College
University of Oxford



A thesis presented for the degree of
Doctor of Philosophy in Sociology

Hilary Term 2026

Word count: 42,135

Thesis Abstract

International migration exhibits persistent empirical regularities in how movement is organised in space and time. These regularities—the spatial distribution of flows, their concentration or dispersion, the reciprocity of exchanges—can be understood as *migration structures*: aggregate manifestations of underlying drivers and mechanisms. Yet, migration research has rarely treated migration structures as analytically meaningful objects. Driver-oriented research has focused on which factors determine migration volumes; pattern-oriented research on how flows are distributed across space. The two literatures rarely intersect, and existing methods offer few tools to connect them.

This thesis argues that migration structure is an essential component of any complete account of human mobility. Across three empirical chapters, it develops the conceptual and methodological apparatus required to study it. The first chapter introduces Hill profiles to examine changes in the diversity of global migration from 1960 to 2020. It shows that the globalisation of international migration proceeded through two independent processes—corridor expansion and flow redistribution—that operated at different historical moments, reconciling contradictory findings in the literature. The second chapter asks whether established migration theories can account for observed migration structures. Developing a simulation-based framework and applying it to intra-European migration flows, it demonstrates that theories achieving high explanatory power may nonetheless fail to reproduce key spatial features of the migration system. The third chapter challenges a foundational assumption in quantitative migration models: the independence of migration decisions. Using a statistical network model applied to European student mobility, it shows that interdependence organises movement beyond what standard predictors explain and that omitting it distorts inference about those predictors.

Taken together, the three chapters make a cumulative case: international migration has structure, structure is empirically tractable, and explaining it requires tools that standard approaches do not provide. This thesis seeks to lay the groundwork for such an agenda.

Acknowledgements

This thesis was made possible by the support of many people and institutions. The ESRC Grand Union Doctoral Training Partnership (GUDTP) and Nuffield College provided funding from the MPhil in Sociology and Demography through to the DPhil, and I am sincerely grateful to both.

My supervisors, Professor Ridhi Kashyap and Dr Charles Rahal, gave me both the freedom and the encouragement to pursue my own ideas. They showed me, by example, what groundbreaking and impactful research in the social sciences can look like. I am grateful for the generosity with which they engaged with my work at every stage, for their patience when my ideas needed time to develop, and for trusting me to find my own way forward.

Professor Per Block's influence on my thinking extends well beyond the chapters of this thesis. Per supervised me during the MPhil and has supported me continuously ever since, inviting me to spend an extended period at the University of Zurich and becoming a collaborator on the third empirical chapter of this thesis. His insistence on thinking sociologically, asking not just whether a model fits but what it actually claims about the world, has shaped the way I approach research in lasting ways. I am deeply thankful for that guidance.

I am grateful to Dr Francesco Rampazzo, Dr Bernie Hogan, and Dr Lindsay Richards, who served as assessors during my Transfer and Confirmation of Status. Their feedback sharpened the thesis at critical moments, and their encouragement gave me confidence that I was on the right track. I am also grateful to the editors of the *European Journal of Population* and three anonymous reviewers for the feedback on the material from the second empirical chapter, which was subsequently published in the journal. I thank Professor Guy J. Abel for his support with the first empirical chapter: he helped construct the underlying dataset, provided detailed feedback, and pushed me to think carefully about what estimated data can and cannot support. I am also thankful to Dr Laura Fortunato for teaching me the importance of rigorous and open science, and for being a model of genuine commitment to diversity, equality, and inclusion both within and beyond academia.

Throughout the DPhil, I benefitted immensely from the vibrant intellectual communities at Nuffield College, the Department of Sociology, and the Leverhulme Centre

for Demographic Science in Oxford, and at the Max Planck Institute for Demographic Research (MPIDR) in Rostock. At MPIDR, I am particularly grateful to Dr Emilio Zagheni, Dr Ebru Şanlıtürk, Dr Christina Westphal, and Dr Heiner Maier for welcoming me into the Doctoral School in Population Health and Data Science and enabling my participation in its courses and Annual Academies. Through that community, I found not only inspiring colleagues but also genuine friends, including Edith, Jasmin, Ji, Jiani, and Liz.

I was fortunate to spend time in Zurich on two occasions: first as part of an Overseas Institutional Visit funded by the ESRC GUDTP and the European Consortium for Sociological Research, and later through an extended collaboration with Per. The team I became part of there made my time in Zurich feel less like a research visit and more like a second home. I am grateful to Marion for her support with the first empirical chapter, and to Linus and Sarah for more brainstorming sessions than any of us can count. A great deal of this thesis was worked out in those conversations.

I thank Filippo for never shying away from lending his keen philosophical eye to my work, and for supporting me through the final stretch of this doctoral journey in more ways than I can easily say here.

Finally, and most importantly, I am grateful to my loved ones scattered across Bologna, Milan, Pievepelago, and Genoa, who have believed in me throughout and never allowed me to forget that there is life beyond the DPhil—and beyond academia. This thesis is for them.

Use of Artificial Intelligence (AI) Tools in the Thesis

Large language model (LLM) tools were used during the preparation of this thesis for two purposes. ChatGPT Edu, as provided by the University of Oxford, was used for language editing and proofreading of written text. GitHub Copilot, integrated within the environment of RStudio, was used to assist with code completion, debugging, and review as part of the empirical analyses. No generative AI tool was used to produce, interpret, or analyse any of the substantive content, arguments, or results presented in this work, and all intellectual contributions are the author's own.

Contents

Thesis Abstract	iii
Acknowledgements	v
List of Tables	xi
List of Figures	xv
List of Acronyms	xvii
1 Introduction	1
1.1 Preview of the Empirical Chapters	3
1.2 Scope of the Thesis	5
2 Theoretical and Methodological Foundations	7
2.1 Migration Systems and Migration Structure	7
2.2 Measuring Structure: Migration Patterns	10
2.2.1 Patterns as Features of Structure	10
2.2.2 Limits of Pattern-Based Approaches	11
2.3 Explaining Structure: Drivers and Mechanisms	12
2.3.1 Migration Drivers	12
2.3.2 Migration Mechanisms and Interdependence	13
2.3.3 Limits of Driver- and Mechanism-Based Approaches	15
2.4 Methodological Background	17
2.4.1 The Network Perspective	17
2.4.2 Simulations and Counterfactual Reasoning	18
2.5 Measuring Migration Flows	19
2.5.1 Flows <i>vs</i> Stocks	19
2.5.2 Limits of Migration Flow Data	20
2.6 Data Sources and Analytical Progression Across the Empirical Chapters	21
3 Diversity and Structural Change in International Migration, 1960–2020	25
Abstract	25
3.1 Introduction	26
3.2 Understanding Migration Diversity	28

3.2.1	Migration Diversity and the Globalisation of Migration	28
3.2.2	The Distributional Problem of Migration Diversity	31
3.2.3	Hill Diversity Profiles as a Unifying Framework	32
3.3	Research Questions and Analytical Strategy	33
3.4	Data and Methods	35
3.4.1	Data	35
3.4.2	Global Hill Diversity Profiles	37
3.4.3	Country-Level Hill Profiles	38
3.4.4	Top- k Dominance	39
3.4.5	Disentangling Corridor Turnover and Flow Redistribution . . .	39
3.4.6	Within- and Between-Country Diversity Decomposition	40
3.5	Results	41
3.5.1	Two Eras of Global Migration Diversity	42
3.5.2	Corridor Turnover <i>vs</i> Flow Redistribution	44
3.5.3	Within- and Between-Country Sources of Diversity	46
3.5.4	Country-Level Trajectories	50
3.6	Discussion	54
4	A Simulation Approach for the Spatial Testing of Migration Theories	63
	Abstract	63
4.1	Introduction	64
4.2	Theoretical and Methodological Background	66
4.2.1	Conceptualising Migration Patterns	67
4.2.2	Conceptualising Migration Drivers	68
4.2.3	Closing the Gap Between Migration Patterns and Drivers	70
4.2.4	Simulation Approaches in Migration Research	71
4.2.5	Simulations for Hypothesis Testing	72
4.3	Proposed Methodology: A Spatial Test for Migration Analysis	74
4.3.1	Step 1: Model Specification	74
4.3.2	Step 2: Simulating Migration Systems	75
4.3.3	Step 3: Extracting Spatial Features	77
4.3.4	Step 4: Statistical Assessment of Spatial Fit	80
4.4	Application: International Migration in Europe, 2002–2021	82
4.4.1	Data	83
4.4.2	Theory Operationalisation	85
4.4.3	Step 1: Regression Results	86
4.4.4	Steps 2–4: Simulation Results	89

4.5	Discussion	96
5	The Interdependence of Migration: How Relational Structures Organise Movement	101
	Abstract	101
5.1	Introduction	102
5.2	Migration Drivers and Relational Interdependence	105
5.2.1	Origin Interdependence	107
5.2.2	Dyadic Interdependence	109
5.2.3	Relational <i>vs</i> Macro-level Drivers	110
5.2.4	Existing Approaches to Modelling Migration Interdependence	112
5.3	Research Questions and Analytical Strategy	114
5.3.1	Inference to Relational Structures	116
5.4	Empirical Case: European Student Mobility	118
5.5	Data	120
5.6	Methods	124
5.6.1	Formulation of the Network Model	124
5.6.2	Formulation of the Interdependence Parameters	125
5.6.3	Model Specification	127
5.6.4	Goodness-of-Fit	128
5.6.5	Relative Importance of Predictors	129
5.6.6	Corridor-Level Model Diagnostics	130
5.7	Results	131
5.7.1	Model Results	131
5.7.2	Measuring Model Performance Across Corridors	137
5.8	Discussion	141
6	Conclusion	149
6.1	Summary of the Empirical Findings	149
6.2	Contributions	151
6.3	Limitations	154
6.4	Areas for Future Research	157
	Bibliography	161
A	Appendix to Chapter 3	185
A.1	Data Availability Statement	185
A.2	List of Countries Included in the Analysis	186

A.3	Distribution of Corridor Volumes, 1960–2020, Pseudo-Bayesian Flow Estimation	194
A.4	Choropleth World Maps of Total Inflows and Outflows, Pseudo-Bayesian Flow Estimation	195
A.5	Chord Diagrams of Directional Migration Flows by World Region, Pseudo-Bayesian Flow Estimation	196
A.6	Hill Diversity of Migration Inflows and Outflows for All Countries, Pseudo-Bayesian Flow Estimation	197
A.7	Country- and Region-Level Contributions to Connectivity and Evenness over Time	213
A.8	Replication with Minimisation–Closed Flow Estimation	215
A.9	Replication with Minimisation–Open Flow Estimation	224
B	Appendix to Chapter 4	233
B.1	Data Availability Statement	233
B.2	List of Countries in the Sample	234
B.3	PPML Regressions with Varying Dependent Variables	235
B.4	Replication with Eurostat Data	238
B.5	Replication with HMigD Subsample	240
B.6	Scatterplots with Observed <i>vs</i> Simulated Flows	241
C	Appendix to Chapter 5	245
C.1	Data Availability Statement	245
C.2	MCMC Convergence Diagnostics of ELMos with Interdependence Parameters (Models 3–4)	246
C.3	Prediction Bias by Corridor for All Models	250
C.4	Corridor-Level Simulation Accuracy for All Models	253

List of Tables

4.1	Step 1: PPML regressions of different migration theories	88
5.1	Summary of full and sampled student populations	123
5.2	Predicting degree and Erasmus student mobility flows using Endogenous Log-linear Models (ELMos)	134
A.1	List of countries included in the analysis, with corresponding ISO 3166-1 codes (alpha-3), continent, and world region labels	186
B.1	List of countries in the sample and corresponding ISO 3166-1 codes (alpha-2 and alpha-3) and UN world region labels	234
B.2	PPML regressions of the baseline model with alternative dependent variables (migration flow quantiles)	235
B.3	PPML regressions of the gravity model with alternative dependent variables (migration flow quantiles)	236
B.4	PPML regressions of the MST model with alternative dependent variables (migration flow quantiles)	237
B.5	Results of PPML regressions using Eurostat data	239
B.6	Results of PPML regressions with HMigD data, excluding years affected by the COVID-19 pandemic (2019–2021)	240

List of Figures

3.1	Hill diversity of the global migration system, 1960–2020	43
3.2	Dominance evolution of corridors in the global migration system . . .	45
3.3	Comparison of changes in global migration diversity through a stable-support analysis	47
3.4	Decomposition of global migration diversity into within- and between-country components	49
3.5	Hill diversity of migration inflows and outflows for selected countries	51
4.1	Diagram of the proposed simulation-based procedure	77
4.2	HMigD estimates of aggregate migration inflows and outflows	84
4.3	Accuracy of two explanatory models in accounting for the aggregate spatial features of the intra-European migration system over time . . .	91
4.4	Distance between observed and simulated inflows in the intra-European migration system by country and over time	93
4.5	Accuracy of two explanatory models in accounting for the migration flows occurring in four exemplary corridors over time	95
5.1	Bilateral flows of degree- and Erasmus-mobile students within 29 European countries in the year 2019	132
5.2	Relative explanatory importance of Model 4 predictors for degree and Erasmus mobility	136
5.3	Prediction bias of Model 4 for all migration corridors in the sample . .	138
5.4	Comparison between explanatory models without interdependence (Model 2) and with interdependence (Model 4)	140
A.1	Distribution of corridor volumes over time, 1960–2020, pseudo-Bayesian flow estimates	194
A.2	Choropleth maps of country-level migration inflows and outflows in the 1960s and 2010s, pseudo-Bayesian flow estimates	195
A.3	Chord diagrams of region-to-region migration flows in the 1960s and 2010s, pseudo-Bayesian flow estimates	196
A.4	Migration diversity in inflows and outflows in East Asia	197
A.5	Migration diversity in inflows and outflows in South-East Asia	198
A.6	Migration diversity in inflows and outflows in South Asia	199

A.7	Migration diversity in inflows and outflows in Central and West Asia	200
A.8	Migration diversity in inflows and outflows in Eastern Europe	201
A.9	Migration diversity in inflows and outflows in Northern Europe	202
A.10	Migration diversity in inflows and outflows in Western Europe	203
A.11	Migration diversity in inflows and outflows in Southern Europe	204
A.12	Migration diversity in inflows and outflows in North Africa	205
A.13	Migration diversity in inflows and outflows in West Africa	206
A.14	Migration diversity in inflows and outflows in East Africa	207
A.15	Migration diversity in inflows and outflows in Middle and South Africa	208
A.16	Migration diversity in inflows and outflows in North America	209
A.17	Migration diversity in inflows and outflows in Central America and the Caribbean	210
A.18	Migration diversity in inflows and outflows in South America	211
A.19	Migration diversity in inflows and outflows in Oceania	212
A.20	Country- and region-level contributions to connectivity and evenness over time	214
A.21	Distribution of corridor volumes over time, 1960–2020, minimisation– closed flow estimates	215
A.22	Choropleth maps of country-level migration inflows and outflows in the 1960s and 2010s, minimisation–closed flow estimates	216
A.23	Chord diagrams of region-to-region migration flows in the 1960s and 2010s, minimisation–closed flow estimates	217
A.24	Hill diversity of the global migration system, 1960–2020, minimisation– closed flow estimates	218
A.25	Evolution of the dominance of the most important corridors in the global migration system, 1960–2020, minimisation–closed flow estimates	219
A.26	Comparison of changes in global migration diversity through a stable- support analysis, minimisation–closed flow estimates	220
A.27	Partitioning of global migration diversity into within- and between- country components, minimisation–closed flow estimates	221
A.28	Hill diversity of migration inflows and outflows for selected countries, 1960–2020, minimisation–closed flow estimates	222
A.29	Country- and region-level contributions to connectivity and evenness over time, minimisation–closed flow estimates	223
A.30	Distribution of corridor volumes over time, 1960–2020, minimisation– open flow estimates	224

A.31	Choropleth maps of country-level migration inflows and outflows in the 1960s and 2010s, minimisation–open flow estimates	225
A.32	Chord diagrams of region-to-region migration flows in the 1960s and 2010s, minimisation–open flow estimates	226
A.33	Hill diversity of the global migration system, 1960–2020, minimisation–open flow estimates	227
A.34	Evolution of the dominance of the most important corridors in the global migration system, 1960–2020, minimisation–open flow estimates	228
A.35	Comparison of changes in global migration diversity through a stable-support analysis, minimisation–open flow estimates	229
A.36	Partitioning of global migration diversity into within- and between-country components, minimisation–open flow estimates	230
A.37	Hill diversity of migration inflows and outflows for selected countries, 1960–2020, minimisation–open flow estimates	231
A.38	Country- and region-level contributions to connectivity and evenness over time, minimisation–open flow estimates	232
B.1	Eurostat estimates of aggregate migration inflows and outflows	238
B.2	Comparison of the gravity-simulated and observed migration flows by region-level corridor over time	241
B.3	Comparison of the MST-simulated and observed migration flows by corridor, year, and time	242
B.4	Comparison of simulated and observed migration flows by corridor, year, and model	243
C.1	Diagnostic plots of Model 3 for degree mobility	246
C.2	Diagnostic plots of Model 4 for degree mobility	247
C.3	Diagnostic plots of Model 3 for Erasmus mobility	248
C.4	Diagnostic plots of Model 4 for Erasmus mobility	249
C.5	Prediction bias of Model 1 for all migration corridors in the sample . .	250
C.6	Prediction bias of Model 2 for all migration corridors in the sample . .	251
C.7	Prediction bias of Model 3 for all migration corridors in the sample . .	252
C.8	Performance comparison across all models for selected corridors	253

List of Acronyms

ABM	Agent-Based Model
ACV	Aggregate Coefficient of Variation
AFMP	Agreement on the Free Movement of Persons
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
CEPII	Centre d'Etudes Prospectives et d'Informations Internationales
CUG	Conditional Uniform Graph
ELMo	Endogenous Log-linear Model
ERGM	Exponential Random Graph Model
EU	European Union
GDP	Gross Domestic Product
HEI	Higher Education Institution
HHI	Herfindahl–Hirschman Index
HMigD	Human Migration Database
ISM	International Student Mobility
MCMC	Markov chain Monte Carlo
MST	Migration Systems Theory
OECD	Organisation for Economic Co-operation and Development
ONS	Office for National Statistics
PPML	Poisson Pseudo-Maximum Likelihood
QAP	Quadratic Assignment Procedure
RESET	Regression Equation Specification Error Test

RI	Relative Importance
SAOM	Stochastic Actor-Oriented Model
SNA	Social Network Analysis
UAE	United Arab Emirates
UIS	Institute for Statistics of the United Nations Educational, Scientific, and Cultural Organisation (UNESCO)
UK	United Kingdom
UN	United Nations
UNDESA	United Nations Department of Economic and Social Affairs
UNESCO	United Nations Educational, Scientific, and Cultural Organisation
US	United States
USSR	Union of Soviet Socialist Republics

1 | Introduction

International migration is not randomly distributed across the globe. Year after year, people move along a relatively stable set of corridors connecting specific origins to specific destinations. Some corridors carry large and persistent flows; others remain dormant or emerge only briefly. Within each origin country, emigrants typically follow one another and cluster in a handful of common destinations, rather than dispersing proportionally to population or opportunity. These regularities represent a ‘fabric’ that is not incidental, but reflects an underlying *structure*: the spatial organisation of movement within migration systems.

Importantly, migration structure is distinct from volume. Two migration systems may exhibit comparable aggregate flows yet differ markedly in how those flows are spatially organised: one concentrated in a few dominant corridors, the other dispersed across many; one characterised by reciprocal exchanges, the other by unidirectional movement. These configurations shape where diasporas form, where remittances flow, and which contexts may become focal points for policy attention. Yet, despite their importance, migration structures have rarely been treated as objects of explanation in their own right. The sociological literature on international migration has developed sophisticated accounts of why some corridors carry more migrants than others, but has paid less attention to what explains the broader architecture of migration systems.

This thesis argues that migration structure is analytically meaningful and worthy of direct investigation. Structure captures properties of migration systems that volume does not, and provides a common object around which otherwise disconnected research traditions—those focused on explaining why people move, and those focused on describing how movement is distributed—can be brought into dialogue. These

considerations raise the guiding question of the thesis: *do our theories of migration, which have proven effective at explaining volume, also explain structure?*

This question arises because the two dominant traditions in quantitative migration research have developed largely in parallel. One tradition, rooted in sociology and economics, has sought to explain migration volumes by identifying the *drivers* and *mechanisms* that determine how many people move along each corridor. Gravity models, Migration Systems Theory (MST), and related theoretical frameworks typically operationalise drivers and mechanisms as regression predictors and evaluate the adequacy of theories by the goodness-of-fit of empirical models. This approach has proven successful: even parsimonious model specifications can account for substantial shares of migration volume. A second tradition, grounded in spatial demography and human geography, has focused on describing the spatial features or *patterns* of migration structure: how concentrated or dispersed flows are in a system, how many corridors are active, or whether the system is becoming more or less interconnected. This literature has produced a rich descriptive toolkit but rarely asked what produces the patterns it documents.

Crucially, the two traditions rarely intersect. Driver-oriented research evaluates theories by their ability to explain variations in flow magnitude, not by whether they reproduce observed spatial configurations. Pattern-oriented research characterises structure but does not connect it to the drivers and mechanisms that may generate it. The result is a gap: we lack tools for asking whether explanations of corridor volumes also account for the system-level properties that emerge when considering those corridors as elements of an integrated system.

Addressing this gap requires not only new concepts but also new methods. This thesis draws on and contributes to the growing field of computational social science, which has introduced several tools for analysing complex social systems and complements traditional regression-based approaches. In particular, Social Network Analysis (SNA) and simulation-based inference provide analytical leverage where standard

methods fall short: they can generate counterfactual systems, recover relational structure from aggregate data, and characterise systemic properties that emerge from the aggregation of individual behaviours.

The three empirical chapters rely on and develop computational methods to engage with migration structure through different epistemic lenses: description (Chapter 3), evaluation (Chapter 4), and inference (Chapter 5).

1.1 Preview of the Empirical Chapters

The first empirical chapter, *Diversity and Structural Change in International Migration, 1960–2020* (Chapter 3), investigates the globalisation of migration hypothesis and examines how the structure of the global migration system changed between 1960 and 2020. Existing research on migration globalisation has reached conflicting conclusions: some studies find that the system has become more diverse over time, while others find that it has become more concentrated. Drawing on new bilateral flow estimates covering 195 countries, the chapter introduces Hill diversity profiles as a framework for analysing the distributional properties of migration systems. The Hill framework reconciles the number of active corridors in the system (connectivity) and the evenness with which flows are distributed across them—two linked dimensions that previous studies analysed separately. This reconciliation proves consequential. The chapter documents that connectivity expanded rapidly through the 1980s before plateauing, while evenness increased primarily from the 1990s onward. The apparent contradiction in the literature thus reflects a genuine finding: globalisation proceeded through two independent structural processes—corridor expansion and flow redistribution—that operated at different historical moments and left different signatures in the spatial organisation of flows.

The second empirical chapter, *A Simulation Approach for the Spatial Testing of Migration Theories* (Chapter 4), shifts from description to evaluation. It asks whether established

migration theories, when operationalised as regression models, can reproduce the spatial configurations of real-world migration systems. The chapter develops a simulation-based framework that generates synthetic migration systems from fitted models and compares their patterns to those of the observed system. Applying this framework to intra-European migration between 2002 and 2021 and testing two prominent theories (the gravity model and MST), the chapter finds that both theories achieve high explanatory power, yet fail to reproduce key features of the system, including the reciprocal nature of European flows and the distinctive migration profiles of Eastern and Northern European countries. The implication is that explaining volume is not the same as explaining structure, and standard theory-testing criteria, focused on models' goodness-of-fit, may overlook spatial accuracy.

The third empirical chapter, *The Interdependence of Migration: How Relational Structures Organise Movement* (Chapter 5), responds to the theoretical limitations documented in Chapter 4. If conventional driver-based models fail to reproduce migration structure, what are they missing? One candidate is *interdependence*: the possibility that migration decisions are not independent of one another, but are linked through social and organisational relations. Standard regression models assume that bilateral flows are conditionally independent given observed predictors—an assumption rarely justified theoretically and potentially consequential for inference. Using a state-of-the-art statistical network model and International Student Mobility flows across 29 European countries, the chapter estimates the effect of interdependence alongside conventional drivers. Comparing degree mobility, where students pursue full degrees abroad, with Erasmus exchanges, where mobility occurs through formal inter-institutional partnerships, the chapter demonstrates that interdependence shapes destination choices in ways that differ systematically across institutional contexts and that standard models cannot capture. Accounting for interdependence also reshapes conclusions about conventional predictors: the effect of geographic distance, for instance, attenuates or reverses once interdependence is modelled explicitly.

1.2 Scope of the Thesis

Taken together, the three empirical chapters make a cumulative case. The first establishes that migration structure exists and changes in patterned ways linked to distinct underlying processes. The second demonstrates that conventional theories, despite their explanatory power, may fail to reproduce that structure. The third offers one account of what they miss: the interdependence that links migration decisions and generates structural regularities not reducible to standard predictors.

Attending to structure—describing it carefully, evaluating whether theories reproduce it, and modelling the relational processes that shape it—yields insights not accessible through approaches focused on migration volume alone. The gap between explaining volume and explaining structure is not a minor technical inconvenience: it reflects a substantive limitation in how migration research has framed its explanatory ambitions. This thesis seeks to make that gap visible and provide tools for addressing it.

The chapters that follow develop the theoretical and methodological foundations for this agenda (Chapter 2), present each empirical study in turn (Chapters 3 to 5), and conclude by reflecting on the contributions and limitations of the thesis as a whole (Chapter 6).

2 | Theoretical and Methodological Foundations

2.1 Migration Systems and Migration Structure

International migration is a complex social phenomenon shaped by forces operating across different levels of analysis. Individual decisions to move are influenced by personal aspirations and constraints, but are also embedded in interpersonal relations, organisational infrastructures, and broader political, economic, and spatial contexts. When aggregated, these individual decisions give rise to structured systems of population movement that exhibit persistent regularities in how mobility is organised across space and time.

One way quantitative migration research has addressed this complexity is through the perspective of *migration systems*. Migration systems conceptualise human movement not as a collection of independent decisions, but as a set of origin–destination linkages that connect places over time (Bakewell, 2024; Bakewell et al., 2012; Boyle et al., 1998; Mabogunje, 1970; Massey et al., 1999). Under this perspective, migration outcomes cannot be understood by analysing sending and receiving contexts in isolation: they must be examined as spatial and relational configurations that evolve together. Origins and destinations are linked through flows whose properties—their distribution, directionality, and reciprocity—constitute analytically meaningful features of the system as a whole.

This thesis adopts the migration systems perspective as its point of departure. Throughout, migration systems are conceptualised as origin–destination flow matrices whose spatial organisation constitutes their *structure*. The term structure is therefore

used differently from the common sociological contrast between ‘structure’ and ‘agency’, where structure denotes macro-level constraints—economic conditions, legal regimes, political contexts—that shape individual migration decisions (Bakewell, 2010). In this thesis, structure is not a set of external forces acting on migrants but an emergent property of the migration system itself: the configuration that arises when individual movements are aggregated across corridors.

Structure, so understood, is a property of the migration system as a whole, not of its individual components, and is independent of flow volume. Two migration systems may exhibit comparable aggregate flows between the same set of countries, yet differ markedly in how those flows are arranged: one highly concentrated, the other dispersed; one characterised by reciprocal exchanges, the other by unidirectional movement. Structure captures these differences and allows us to ask why migration systems are organised in particular ways.

The claim that migration structure is analytically meaningful requires justification. Why should researchers consider the entire migration system and attend to how flows are organised within it, rather than simply explain how many people move along each corridor? Three considerations motivate this attention. First, structure captures properties of migration systems that volume does not. Differences in spatial organisation matter for policy, integration, and the lived experience of migration, yet are invisible to analyses that focus on volume alone. Structure makes such differences legible (de Haas et al., 2022; DeWaard & Ha, 2019). Second, structure is consequential. The spatial organisation of a migration system shapes the distribution of migration’s effects: where remittances flow, where diasporas form, where labour markets tighten or slacken. A highly concentrated system channels these effects into a small number of corridors; a dispersed system distributes them more broadly. Understanding how structure is produced and how it changes is therefore a prerequisite for understanding the broader consequences of migration (Bakewell, 2014). Third, structure provides

a common object around which otherwise disconnected research traditions can be reconciled. Demographic and geographic research has developed sophisticated tools for describing patterns—the observable configurations that structure takes—but has often stopped short of asking what produces them. Sociological and economic research has identified the drivers and mechanisms that trigger and sustain migration, but has rarely evaluated whether these accounts reproduce observed spatial configurations. Treating structure as the integrating concept makes it possible to bring these traditions together: to ask not only whether drivers explain volume, but whether they reproduce structure; to investigate not only which mechanisms are operating, but whether their signatures are visible in the spatial organisation of movement.

This systemic orientation aligns the thesis with broader computational advancements in the social sciences (Edelmann et al., 2020; Kashyap et al., 2023; Lazer et al., 2009, 2020). Where traditional approaches model migration as a collection of independent bilateral flows, each explained by the attributes of its origin and destination, the systems perspective treats the full configuration of flows as the unit of analysis, and its properties as emergent features that cannot be reduced to corridor-level characteristics. Computational methods such as Social Network Analysis (SNA) and simulations are well suited to this task: they can simulate entire systems under alternative assumptions, estimate parameters that capture interdependence between flows, and characterise distributions rather than point estimates (Bijak et al., 2018; Bilecen et al., 2018; Drouhot et al., 2023; Stadtfeld & Amati, 2021).

The empirical chapters deploy computational methods in sequence, moving from a distributional analysis of global corridor diversity (Chapter 3), through simulation-based assessments of whether migration theories reproduce observed spatial configurations (Chapter 4), to statistical network models that estimate interdependence as an organising principle of human mobility (Chapter 5).

2.2 Measuring Structure: Migration Patterns

2.2.1 Patterns as Features of Structure

Migration patterns are the empirical means through which structure becomes visible and an object of description. Where structure refers to the underlying spatial organisation of flows, patterns are the measurable expressions of that organisation: how concentrated or dispersed flows are, how reciprocal or asymmetric, or how connected the system is.

A wide range of descriptive tools has been developed in spatial demography, human geography, population ecology, and network science to characterise migration patterns. Among these, indices of migration connectivity describe the density and richness of corridors within a system (Abel et al., 2021; Charles-Edwards et al., 2025; Davis et al., 2013; Ha et al., 2025). Indices of migration evenness or concentration, also known as spatial focusing, capture how evenly or unevenly flows are distributed across these corridors (Bell et al., 2002; Czaika & de Haas, 2014; Plane & Mulligan, 1997; Rogers & Raymer, 1998; Rogers & Sweeney, 1998). Indices of migration reciprocity capture the balance of bilateral exchange between pairs of locations in the system (Huang & Butts, 2023; Leal, 2021; Leal & Harder, 2021, 2023; McMillan, 2024).

The relationship between structure and patterns is analogous to that between a distribution and its summary statistics. Structure is the full configuration of flows across corridors; patterns are the specific features of that configuration that researchers choose to measure. A single structure can exhibit different patterns depending on which indices are applied, and different indices may rank systems differently depending on which part of the distribution they emphasise. As discussed in greater detail in Chapter 3, this is a reflection of the multidimensional nature of migration structure itself.

2.2.2 Limits of Pattern-Based Approaches

Descriptive approaches to migration patterns have generated valuable insights, but they are not typically concerned with explaining why certain patterns emerge. Indices of connectivity, evenness, and reciprocity characterise the spatial organisation of flows but do not identify the drivers or mechanisms that produce it. Pattern-oriented research documents *what* migration structure looks like; it does not address *why* it looks the way it does. This explanatory gap motivates this thesis, which turns to driver- and mechanism-based accounts to explain how migration structures come to take the forms they do.

Throughout the thesis, patterns are used to measure and describe migration structure. In Chapter 3, migration structures are the central *objects of description*: the chapter relies on diversity patterns—specifically, measures of how many corridors carry flows (connectivity) and how evenly flows are distributed across corridors (evenness)—to document how the distributional shape of the global migration system has evolved over time. In Chapter 4, migration structures are *objects of evaluation*: the chapter asks whether the patterns observed in the intra-European migration system—its concentration, reciprocity, and country-specific migration profiles—are reproduced by empirical models that operationalise established migration theories.¹ In Chapter 5, migration structures function as *empirical signatures*: this chapter treats patterns as the observable imprints that migration mechanisms leave on aggregate flows, using them to detect the presence of processes that cannot be measured directly. In all three cases, migration structure is treated not as statistical ‘noise’ or residuals to be controlled away, but as a substantive manifestation of the migration process, and one that any adequate empirical account must be able to characterise and reproduce.

¹The concentration measures used in Chapter 4 are complements of the evenness measures used in Chapter 3: high concentration corresponds to low evenness, and vice versa. However, whereas Chapter 3 characterises evenness along a continuous spectrum, Chapter 4 relies on single-point indices that each capture a specific location on that spectrum. The diversity framework in Chapter 3 reveals that these indices weight different parts of the distribution differently; the single-point measures in Chapter 4 are chosen for ease of assessment in the theory evaluation and comparability with prior literature.

2.3 Explaining Structure: Drivers and Mechanisms

2.3.1 Migration Drivers

Migration drivers refer to the conditions and forces that influence whether, when, and where migration occurs. Drivers describe what exists in a given context: resources, opportunities, constraints, and connections that make migration possible, attractive, or costly. They are conditions rather than processes, and as such they specify migration potential without necessarily explaining how migration unfolds after that potential is realised (Czaika & Reinprecht, 2022; de Haas, 2021; Van Hear et al., 2018).

Drivers are conventionally organised across three analytical levels. At the micro level, they include individual aspirations, capabilities, and life-course characteristics (de Haas, 2021). At the meso level, they encompass interpersonal networks, organisational infrastructures, and institutional linkages connecting origins and destinations (Faist, 1997, 2000; Xiang & Lindquist, 2014). At the macro level, they refer to contextual conditions such as economic differentials, labour demand, political stability, migration regimes, and geographic distance (Mayda, 2010). This tripartite distinction has become a fundamental organising principle in migration studies and a means of differentiating between competing theoretical accounts (Czaika & Reinprecht, 2022; de Haas, 2021; Massey et al., 1993; Van Hear et al., 2018).

A large share of the quantitative literature on international migration operationalises drivers within regression-based models of country-to-country flows. In this tradition, migration theories are translated into sets of explanatory variables, and empirical support for a theory is assessed through statistical significance, effect sizes, or overall goodness-of-fit (Garip, 2012). Gravity models represent one of the earliest and most influential examples of this approach, positing that migration between two locations increases with population size and decreases with geographic distance (Beine et al., 2016; Lewer & Van den Berg, 2008; Poot et al., 2016; Ramos, 2016; Ravenstein,

1885, 1889). More elaborate frameworks—including theories of cumulative causation and Migration Systems Theory (MST)—extend gravity-type specifications to include historical, institutional, and relational predictors such as shared language, colonial ties, and prior migrant stocks (DeWaard et al., 2012; Leal, 2021; Leal & Harder, 2021, 2023; Massey, 1990; Massey et al., 1993).

Driver-based models have proved remarkably successful at explaining variation in international migration volumes. Even parsimonious specifications such as the gravity model routinely achieve high explanatory power, indicating that a substantial share of corridor-level variation can be accounted for by a low-dimensional set of predictors (Termote, 2002). Nonetheless, the evidence suggests that drivers are often better suited to explaining the *initiation* of migration rather than its *continuation* (Riosmena, 2024). Conditions that prompt the first movers along a corridor may weaken or reverse over time, yet flows often persist. This pattern points to the limits of driver-based accounts: they specify when and why migration becomes possible, but not necessarily how it becomes self-sustaining once underway (Arango, 2000; Bakewell, Engbersen et al., 2016; de Haas, 2010; Massey, 1990).

2.3.2 Migration Mechanisms and Interdependence

A substantial literature has developed in response to these limitations and explains why migration might persist after the conditions that initiated it disappear, why flows concentrate in particular destinations rather than dispersing in proportion to the availability of opportunities, and why migration systems exhibit path dependence, self-reinforcement, and feedback that cannot be reduced to contextual factors.

This literature typically invokes *mechanisms*—processes that link individual decisions to aggregate outcomes through unobservable pathways—to account for these regularities. While drivers describe the conditions under which migration becomes possible, mechanisms are the relational processes through which migration potential is shaped, channelled, and reproduced over time. For instance, the presence of other

actors in one's social network is a driver, while the peer influence operating through those actors is a mechanism. A formal institutional partnership is a driver, while the organisational channelling through that partnership is a mechanism. In this sense, mechanisms are not additional explanatory variables, nor do they simply raise or lower migration volumes: they specify how migration unfolds once drivers are in place.

This conceptualisation aligns with the broader programme of analytical sociology, which treats mechanisms as the processes through which individual actions produce aggregate outcomes that cannot be deduced from a theory of action alone (Åberg, 2000; Granovetter, 1985; Stadtfeld & Amati, 2021). In this framework, mechanisms occupy a distinct explanatory step: they specify how actors link their decisions into reproducible constellations, the collective consequences of which constitute the emergent social phenomena to be explained (Hedström, 2005; Hedström & Bearman, 2009; Hedström & Ylikoski, 2010; Schmid, 2011; Ylikoski, 2011).

Key migration mechanisms identified in the literature include peer influence and chain migration, whereby established migrants facilitate subsequent moves by others from the same origin (Garip & Asad, 2016; Lindstrom & López Ramírez, 2010; MacDonald & MacDonald, 1964); cumulative causation, whereby chain migration becomes self-sustaining after flows reach a critical mass (Massey, 1990; Massey et al., 1993); migrant competition, whereby the attractiveness of a destination depends on the availability of alternatives (Bertoli & Fernández-Huertas Moraga, 2013; Stouffer, 1960); and organisational channelling, whereby organisations such as firms, universities, and recruitment agencies coordinate movement across individuals and places (Cranston et al., 2018; Faist, 1997, 2000; Findlay & Li, 1998; Xiang & Lindquist, 2014).

The defining feature of these mechanisms, as conceptualised in this thesis, is that they generate *interdependence*: migration decisions are linked through social and organisational relations, such that one person's choice affects the decision environment faced by others. Interdependence has long been recognised as central in analytical sociology and migration theory, often through the concept of 'migration feedback'

(Bakewell, Engbersen et al., 2016; de Haas, 2010). Nonetheless, it has rarely been treated as a general organising principle that links individual decisions into structured systems of mobility. This thesis develops that connection in Chapter 5, where interdependence is studied as the empirical manifestation of mechanisms and operationalised within a statistical network model.

2.3.3 Limits of Driver- and Mechanism-Based Approaches

Three limitations of driver- and mechanism-based approaches motivate the analytical strategy pursued in this thesis. The first concerns the scope of driver-based explanations. While drivers have proved effective at explaining variation in migration volumes, it remains an open question whether they account equally well for the spatial organisation of migration systems—that is, for structure. A model may correctly predict that large flows occur between neighbouring countries, yet fail to distinguish whether this describes migration from Mexico to the United States (US) or from Austria to Germany—two corridors with similar dyadic characteristics but embedded in migration systems with very different structures. Additionally, models may not account for the system-level patterns emerging from the specific spatial distribution of flows, such as connectivity, evenness, or reciprocity. Put differently, goodness-of-fit at the corridor level may not guarantee spatial accuracy at the system level. This thesis advances the claim that driver-based models may explain volume better than structure, and tests it directly in Chapter 4.

A related limitation is that standard driver-based models assume bilateral flows are conditionally independent given observed predictors. Under this assumption, flows from any origin to a destination are modelled as independent of migration between other origin–destination pairs, once relevant covariates are controlled for. This assumption is rarely justified theoretically: rather, it reflects a modelling convenience inherited from the econometric and gravity literature (Aksoy & Yıldırım, 2024; Batista & McKenzie, 2023; Beine et al., 2025). As discussed in Chapter 5, migration theory

provides strong reasons to expect interdependence between flows, and when such interdependence exists but is not modelled, coefficient estimates may be biased and spatial properties poorly reproduced.

The third limitation concerns the measurement of mechanisms. Many of the key mechanisms discussed above—peer influence, chain migration, cumulative causation, migrant competition, and organisational channelling—are difficult to measure directly, especially at scale (Bakewell, Kubal & Pereira, 2016). Capturing these dynamics typically requires individual-level data on social ties and organisational information about institutional coordination. Such data are virtually impossible to obtain systematically and comprehensively, particularly for international migration. Chapter 5 addresses these measurement difficulties by treating migration structure as the observable manifestation of underlying mechanisms. Migration mechanisms generate empirical regularities in the spatial organisation of movement, leaving ‘imprints’ on aggregate flows: for instance, peer influence, chain migration, and cumulative causation produce concentration, while migrant competition produces evenness. By estimating these structural features as parameters, it becomes possible to detect the presence of mechanisms that cannot be observed or measured directly.

This inferential strategy reflects a broader epistemological position within analytical sociology. Research in this tradition has long argued that what is observed in aggregate is the result of latent mechanisms that may operate at different analytical levels (Hedström, 2005; Hedström & Bearman, 2009; Ylikoski, 2011). Isolating and estimating structural parameters from aggregate data thus helps recover the composite trace of these latent mechanisms. This strategy is a viable and promising alternative when direct mechanistic identification is not feasible, as it satisfies the criterion of ‘explanatory relevance’ (Ylikoski, 2011, p. 170): if changing the *explanans* (for instance by modelling interdependence rather than assuming independence) changes conclusions about the *explanandum*, then the *explanans* is explanatorily relevant regardless of whether the underlying causal process is fully specified.

2.4 Methodological Background

Addressing the limitations identified above requires methods that go beyond the regression-based toolkit that dominates quantitative migration research. The thesis draws on two computational traditions that provide the necessary analytical leverage: SNA and simulation-based inference. SNA models treat the full configuration of migration flows as the unit of analysis, linking individuals to the wider system. Simulation-based approaches generate synthetic systems under specified theoretical assumptions, enabling direct comparison between the spatial properties of model-implied and observed migration systems. Both traditions share a systemic orientation: they analyse migration not as a collection of independent decisions but as an emergent configuration whose properties—concentration, connectivity, reciprocity, interdependence—arise from the aggregation of individual moves.

2.4.1 The Network Perspective

Migration networks provide a conceptualisation analogous to migration systems. In migration networks, locations are the *nodes* and the flows moving between them are the *weighted, directed ties*. This representation—which captures ‘macronetworks’ of migration flows between places, distinct from ‘micronetworks’ of interpersonal ties among migrants (Danchev & Porter, 2021)—makes explicit the idea that migration systems are relational: they describe not only properties of individual migrants and locations, but also the connections between them. The network lens has become increasingly common in migration research (Bilecen et al., 2018; Drouhot et al., 2023), in part because network-analytic tools can characterise the topology of the system in ways that complement traditional migration indices, for instance through measures of network connectivity and reciprocity (Akbari, 2021; Danchev & Porter, 2018; Davis et al., 2013; Huang & Butts, 2024, 2023; Leal, 2021; Leal & Harder, 2021, 2023; McMillan, 2024; Tranos et al., 2015).

For the purposes of this thesis, the network representation is foundational because it foregrounds the possibility of *interdependence* between flows. In a network, the presence or magnitude of one tie may be related to the presence or magnitude of other ties. Standard regression models treat each tie as conditionally independent; statistical network models allow for the possibility that ties are interdependent. Even when the network representation is not explicitly invoked, it informs all three empirical chapters: Chapter 3 analyses the distributional properties of the global migration network, Chapter 4 evaluates whether simulated networks reproduce observed spatial configurations, and Chapter 5 uses statistical network models that capture interdependence between migration decisions.

2.4.2 Simulations and Counterfactual Reasoning

The other common methodological thread across the three chapters is the reliance on simulations and counterfactual reasoning as the primary tools for inference. Simulation-based approaches have become increasingly central to sociological and demographic inquiry, enabling researchers to generate synthetic populations, test theoretical mechanisms, and explore scenarios that cannot be observed directly (Bijak, 2022; Drouhot et al., 2023; Kashyap & Villavicencio, 2016). In migration research specifically, simulations have been used to model how migration routes emerge and stabilise (Bijak, 2022) and how internal mobility responds to external shocks (Huang & Butts, 2024). The present thesis extends this tradition by using simulations not only to generate counterfactual scenarios but also to evaluate the spatial accuracy of migration theories and to estimate interdependence parameters via Markov chain Monte Carlo (MCMC).²

²The use of simulations throughout the thesis reflects an understanding of prediction and inference as *complementary* rather than *competing* modes of analysis. Recent methodological work has argued that predictive accuracy can serve as a diagnostic for explanatory adequacy, with the two modes reinforcing one another (Hofman et al., 2021). This contrasts with earlier framings that sharply distinguished ‘explanatory modelling’ from ‘predictive modelling’ and treated them as requiring fundamentally different analytical choices (Shmueli, 2010). This thesis deliberately moves between the two registers, using predictive performance to evaluate explanatory theories and inferential models to recover the parameters that prediction alone cannot identify.

In Chapter 3, counterfactual analysis via simulation is used to isolate the contributions of two distinct structural processes—corridor turnover and flow redistribution—to the diversity of the global migration system. This decomposition makes visible how much of the observed longitudinal changes in the flow distribution is due to corridors entering and exiting the system, or due to a progressive reallocation of migratory mass away from dominant corridors.

In Chapter 4, simulations are used to generate synthetic migration systems from fitted regression models. The procedure draws from the flow distribution implied by each model’s estimated parameters and produces an ensemble of counterfactual systems that represent how migration *would* be organised if the theory were correct. Comparing the spatial properties of the simulated systems to those of the observed system reveals whether a theory that explains volume can also reproduce structure.

In Chapter 5, simulations are used to estimate the interdependence parameters via MCMC methods, and later to assess model fit by comparing observed and model-derived migration systems, adopting the strategy of Chapter 4. In this case, the counterfactual contrast asks: how would the distribution of flows differ if migration decisions were independent *vs* interdependent? By comparing models with and without interdependence parameters, the chapter quantifies to what extent interdependence contributes to the spatial organisation of flows.

2.5 Measuring Migration Flows

2.5.1 Flows *vs* Stocks

This thesis studies international migration through country-to-country, bilateral *flows*: the number of people moving between an origin and a destination over a defined period. Conversely, migrant *stocks* measure the number of foreign-born or foreign-national residents in a country at a given point in time.

The choice of flows over stocks is motivated by the thesis's focus on the spatial organisation of migration systems. Flows capture directionality and temporal dynamics in ways that stocks cannot: a stock matrix records the cumulative outcome of decades of migration, mortality, and return, conflating current movement with historical settlement. For this reason, it may obscure whether a bilateral connection reflects ongoing mobility or a legacy of movement that has since ceased (Ha et al., 2025). Flows, by contrast, register active exchanges between places within a defined period. They are the natural unit of analysis for studying structural change—how corridors emerge, grow, contract, or redirect over time—and for examining features such as reciprocity, which requires information about the specific direction of movement within corridors.

The analytical object throughout the thesis is therefore the migration flow matrix and its spatial organisation. Stocks may serve as inputs (for instance, to the estimation of flows in Chapter 3, or as lagged predictors of current flows in Chapter 4) but structure is defined in terms of how flows are distributed across corridors at a given time or how that distribution changes between periods. The network representation reinforces this priority: flows correspond to the weighted, directed ties connecting nodes, whereas stocks would represent attributes of each origin–destination dyad.

2.5.2 Limits of Migration Flow Data

The analytical advantages of migration flows come at a cost: bilateral migration flow data are considerably more difficult to obtain and compare across countries than stock data. Countries differ in how they define an international migrant—by citizenship, country of birth, duration of stay, or registration status—and these definitions shape who is counted and who is missed (Yildiz & Abel, 2021). Registration infrastructures vary considerably, too: countries with population registers produce more complete and timely flow data than those relying on surveys, border records, or residence permits (Daňko et al., 2024). Further, many forms of migration, including short-term, circular, return, and irregular movement, are systematically underrecorded (Fransen & de Haas,

2022). Flow data are also typically available only at annual or multi-year intervals, obscuring within-period dynamics (Fiorio et al., 2021).

These limitations interact with the study of migration structure in specific ways. Underreporting of return and circular flows, for instance, may affect the measurement of reciprocity, potentially biasing estimates downward in systems where such mobility is prevalent, as in the intra-European case considered in Chapter 4. Further, differences in migrant definitions across countries may introduce measurement heterogeneity and distort cross-national comparisons. Recent efforts to develop estimation strategies that produce more consistent flow measures are promising (cf Daňko et al., 2024), but the use of flow estimates rather than observed, registered counts still introduces additional uncertainty, as discussed in Chapters 3 and 4.

These considerations do not invalidate the use of flow data in international migration research, but they require that empirical analyses confront measurement uncertainty directly. Where flows are not directly observed, the estimation process itself introduces assumptions about the relationship between stocks and flows, about how to reconcile inconsistent sources, and about how to handle missing data. The empirical chapters address these concerns through robustness checks, alternative estimation strategies, and explicit acknowledgments of inferential limits.

2.6 Data Sources and Analytical Progression Across the Empirical Chapters

The three empirical chapters build on the concepts and considerations illustrated above in a cumulative sequence. Each chapter engages with the structure of migration, but does so through a different epistemic lens, analytical ambition, and data environment. The progression from estimated global flows to harmonised regional flows to admin-

istrative counts represents a deliberate tightening of measurement precision as the thesis moves from description to evaluation to explanation.

To describe how the structure of the global migration system has changed over time, Chapter 3 develops new global bilateral migration flow estimates covering 195 countries over six ten-year periods. This is done by reconstructing flows from bilateral migrant stocks released by the World Bank (2011; 2023) using demographic accounting methods developed by Abel (2018), Abel and Cohen (2019, 2022), Abel and Sander (2014), and Azose and Raftery (2019). This data environment is appropriate for the chapter's descriptive aims, where geographic and temporal coverage matter more than the precision of individual corridor estimates. Moreover, because the analytical focus is on the shape of the flow distribution rather than on volume, small differences in estimated corridor counts are less consequential than they would be in a study of migration volumes: what matters is whether estimation would affect the distributional properties that the chapter seeks to characterise. The chapter relies on three different flow estimation methods, where one serves as the primary specification and the other two as robustness checks of the findings.

Chapter 4 moves from describing structure to evaluating it. It uses annual harmonised bilateral migration flow data for 31 European countries from the Human Migration Database (HMigD; Dańko, 2023; Dańko et al., 2024), covering the period 2002–2021. These data integrate information from multiple administrative sources through Bayesian estimation methods, and provide finer temporal resolution over a more recent period than the global estimates of Chapter 3. This choice was motivated by the evaluative aims of the chapter: testing whether migration theories reproduce observed structure requires granular data and consistent cross-country migrant definitions, while still retaining sufficient temporal and geographical scope to assess theory performance across different contexts.

Chapter 5 completes this progression by modelling migration structure directly: specifically, the residual structure that conventional drivers leave unexplained. It

uses register-based administrative counts of International Student Mobility across 29 European countries in 2019: data on degree mobility from a joint UNESCO-OECD-Eurostat data collection (UNESCO Institute for Statistics, 2023) and data on Erasmus mobility from administrative records of the European Commission (European Commission, Directorate-General for Education, Youth, Sport and Culture, 2021). These data sources provide registered counts rather than estimates, guaranteeing the strongest measurement foundation of the three chapters, at the cost of a narrower scope. This precision is appropriate for the chapter's inferential goal, which requires reliable bilateral flow data from different institutional settings to estimate the interdependence parameters in the statistical network model.

The analytical progression across the three chapters is intentional: from a descriptive account that reveals the importance of structure (Chapter 3), to the identification of what conventional models fail to reproduce (Chapter 4), to a promising solution to such shortcomings (Chapter 5). The data choices align with this progression: estimated flows with global coverage suffice for documenting broad structural trends; harmonised regional flows permit the finer-grained spatial comparisons required for theory evaluation; and directly observed administrative counts provide the measurement precision necessary for estimating interdependence parameters. Each chapter relies on data appropriate to its analytical goals: as the inferential ambition of the thesis increases, so do the data and measurement standards required to support it.

3 | Diversity and Structural Change in International Migration, 1960–2020

Abstract

Whether international migration has become more diverse over time is a central question in the migration globalisation literature, and one that different studies have answered inconsistently. I argue that these inconsistencies are not measurement or data artefacts but a substantive finding: existing diversity indices capture different parts of the distribution of flows across origin–destination corridors, and those parts have moved in opposite directions over time. Using Hill profiles, which nest different diversity indices as special cases of a single continuous function, and new bilateral flow estimates for 195 countries from 1960 to 2020, I identify two distinct eras of global migration diversity: expanding corridor connectivity through the 1980s, followed by growing evenness from the 1990s onwards. Counterfactual analysis shows that these eras reflect separate processes: corridor expansion in earlier periods, and reallocation of migration mass across persistent corridors in more recent ones. A within- and between-country decomposition reveals that post-1990 changes in migration diversity have been driven by countries individually diversifying their immigration and emigration portfolios, while converging on increasingly similar dominant partners. The apparent paradox of migration globalisation thus dissolves once migration diversity is understood as a distributional property rather than a single summary measure.

3.1 Introduction

Over the past six decades, international migration has grown substantially in both volume and geographic reach, drawing sustained scholarly and policy attention to its changing *diversity*: how many country pairs are connected, how migration flows are distributed across corridors, and whether this distribution has become more or less concentrated over time. Questions of migration diversity are central to debates about the globalisation of migration, and the answers have broad consequences for how we understand the governance of international movement and the integration of societies (Czaika & de Haas, 2014; de Haas et al., 2022; Ha et al., 2025; Massey et al., 1999).

Existing research has relied on two main operationalisations of migration diversity. Some studies focus on connectivity, counting all active origin–destination corridors and treating them as equivalent regardless of how many migrants they carry, and find that the global migration system expanded rapidly through the 1970s (Davis et al., 2013) before plateauing or contracting (Ha et al., 2025). Other studies focus on how evenly migration volume is distributed across corridors, weighting them by their share of total migration, and find that evenness has increased over time (Czaika & de Haas, 2014; Fitter et al., 2024). Both approaches seek to measure ‘migration diversity’, yet they produce inconsistent assessments of whether the global migration system has become more or less diverse over time.

This inconsistency is not merely technical. It reflects a deeper ambiguity in what migration diversity truly means. Corridors differ enormously in the number of migrants they carry: a handful of highly popular routes account for most international movement, while the majority of corridors carry relatively few migrants. Any diversity index must take a position on how much large corridors should count relative to small ones, and different indices take different positions. A connectivity index weights all corridors equally; an evenness index gives greater weight to dominant corridors. Because they emphasise different parts of the corridor distribution, these indices can

move independently: whether they do is an empirical question that the literature has not directly addressed.

This chapter introduces Hill profiles as a framework for studying migration diversity that makes these weighting choices explicit. Developed in ecology to unify indices of species diversity, Hill profiles parameterise the weighting of categories in a system through a single sensitivity metric, transforming competing diversity indices into points on a continuous spectrum (Hill, 1973). The resulting diversity profile—a curve traced out as sensitivity varies—reveals the full distributional shape of the system, rather than collapsing it to a single number. Applied to international migration, the Hill diversity framework varies the contribution of small *vs* dominant corridors, reconciling connectivity and evenness and making their tension visible and analytically tractable.

Applying this framework to new estimates of bilateral migration flows in 195 countries from 1960 to 2020, I show that connectivity and evenness are empirically decoupled and have dominated the global migration system at different times. Connectivity peaked in the 1980s and subsequently declined; evenness declined slightly through the 1970s before rising steadily from the 1990s onwards. Counterfactual analysis shows that global diversity change was driven by two distinct processes: in the earlier periods, by the emergence of new migration connections (corridor turnover); in more recent periods, by a reallocation of migrants from dominant toward smaller and medium-sized corridors (flow redistribution). I further decompose inflow and outflow diversity into within- and between-country components to distinguish whether global diversity change reflects countries developing more balanced portfolios of linked origins and destinations (within-country diversification) or becoming more distinct from one another in their dominant connections (between-country differentiation). Together, the results define two distinct eras of migration diversity: an era of corridor expansion and between-country differentiation through the 1980s, followed by an era of flow redistribution and within-country diversification from the 1990s onwards.

The chapter makes three contributions. Conceptually, it reframes migration diversity as a distributional property, showing that connectivity and evenness are complementary views of the same object rather than competing measures. Methodologically, it introduces Hill diversity profiles and a within–between diversity decomposition to migration research. Substantively, it demonstrates that migration diversification has proceeded through two distinct structural processes, corridor turnover and flow redistribution, and that the post-1990 rise in diversity reflects within-country diversification rather than increasing differentiation between countries. These findings clarify a persistent puzzle in the literature: studies have reached different conclusions about migration globalisation because they have examined different parts of the same corridor distribution.

The remainder of the chapter proceeds as follows. The next section develops the distributional framework in greater depth and situates Hill profiles within the broader literature on migration globalisation and diversity. I then describe the data and methods before presenting the empirical results, moving from global patterns to country-level trajectories. I conclude by discussing the implications and limitations of the findings.

3.2 Understanding Migration Diversity

3.2.1 Migration Diversity and the Globalisation of Migration

International migration connects world countries through sustained bilateral flows, giving rise to empirical regularities in how human movement is organised in space and time. In this sense, international migration can be understood as a *system*: a configuration of countries and the flows that connect them (Bakewell, 2014; Bakewell, 2024; Boyle et al., 1998; de Haas et al., 2022; DeWaard & Ha, 2019; Mabogunje, 1970).¹

¹This conceptualisation is analogous to that of macroscale ‘migration networks’, where countries (the ‘nodes’) are connected by migration flows (the ‘ties’; Danchev & Porter, 2021).

To analyse these configurations, also known as migration patterns, scholars have developed summary indices that compress origin–destination flow matrices into single scalars. Changes in these indices are interpretable as evidence of structural change in the migration system: shifts in which corridors are active, how flows are distributed across them, and how the system relates to the geographic and institutional landscape in which it is embedded (Bakewell, 2014; DeWaard & Ha, 2019; Rogers & Raymer, 1998).

Of these structural changes, the globalisation of migration has garnered the most scientific scrutiny. The globalisation of migration refers to the process by which international migration expands along three dimensions: *intensity*, or the propensity of people to move across borders relative to the population at risk; *distance*, or the average geographic reach of international moves; and *diversity*, or the breadth and balance of flows across corridors (Czaika & de Haas, 2014; Ha et al., 2025). The first two dimensions have been measured consistently across studies and generated broadly comparable findings: intensity has remained stable as a share of world population, and distance has increased modestly since the 1960s (Czaika & de Haas, 2014; de Haas et al., 2022; Fitter et al., 2024; Ha et al., 2025). Diversity, by contrast, has proven more difficult to operationalise, and the literature has approached it from two parallel perspectives.

The first perspective operationalises migration diversity as *connectivity*: how many countries are involved in the global migration system and how many bilateral corridors are active. Connectivity is most commonly measured as the count of corridors for which a positive migration flow is observed, or as the ratio of active corridors to the maximum possible number in a migration system of a given size (Abel et al., 2021; Charles-Edwards et al., 2025; Davis et al., 2013; Ha et al., 2025). Both formulations treat all corridors as equivalent regardless of migration volume: a corridor carrying a handful of migrants contributes to connectivity in the same way as one carrying millions.

Using World Bank migrant stock data for 226 countries between 1960 and 2000, for instance, Davis et al. (2013) document a sustained expansion in connectivity and

interpret this trend as evidence of growing diversity and integration of the global migration system. Looking at more recent periods and drawing on UN-derived flow estimates, rather than stocks, Ha et al. (2025) find that connectivity showed no sharp increase between 1990 and 2020, suggesting that diversity through increased connectivity had largely plateaued by the end of the 20th century.

The second perspective operationalises migration diversity as *evenness*: whether flows are spread broadly across active corridors or concentrated in a small number of highly popular ones. Evenness indices are sensitive not only to how many corridors are in use but also to the relative size of flows moving along them. In practice, evenness indices are most commonly based on the Herfindahl–Hirschman Index (HHI), also known as ‘spatial focusing’ in the internal migration literature (Bell et al., 2002; Plane & Mulligan, 1997; Rogers & Raymer, 1998; Rogers & Sweeney, 1998), and computed as the sum of squared flow shares across all active corridors. Because squaring the shares gives disproportionate weight to the largest corridors, the HHI is high when migration is concentrated in a few dominant corridors and low when flows are spread evenly. Studies of migration globalisation typically use the complement of this measure (one minus the HHI), so that higher values indicate greater evenness, in what is also known as ‘migration spread’ (Czaika & de Haas, 2014).²

Using this measure and World Bank stock data for 226 countries, Czaika and de Haas (2014) document increasing global diversity from 1960 to 2000. They additionally compute inflow and outflow diversity both at the global and country level, and find that immigration portfolios (ie sets of countries of origin of immigrants) became more diverse over the period (increasing immigration spread), even though the pool of popular destination countries had shrunk (decreasing emigration spread). The authors refer to this as an ‘essential paradox’ of migration: while global migration diversity increased, country-level outflow diversity decreased. Fitter et al. (2024)

²The Shannon entropy provides an alternative operationalisation of evenness that is less dominated by the largest corridors: rather than squaring the flow shares, it weights each corridor proportionally to its share. However, Shannon-based diversity measures are less common in migration research.

extend this analysis from 1990 to 2020 using UN-derived flow estimates, finding that the global trend toward increased evenness continued after 2000, but at a slower pace. They additionally document a reversal in the divergence between inflow and outflow diversity, showing that, after 2000, outflow diversity began to increase as well.

3.2.2 The Distributional Problem of Migration Diversity

The inconsistencies in prior evidence of migration diversity stem from a general problem that arises whenever one tries to summarise a system whose elements are distributed across categories of unequal size. In the global migration system, the elements are individual migrants and the categories are origin–destination corridors. The distribution of migrant mass is highly skewed: the vast majority of migrants are moving along a small number of corridors, while most other corridors carry comparatively little traffic. Summarising this distribution as a single number requires taking a stance on how much the dominant corridors should count relative to peripheral ones: this stance, whether explicit or not, is what distinguishes the diversity indices used in the literature.

To make the problem concrete, consider a hypothetical migration system being analysed in two periods, with the total number of migrants being constant. In period t , the system has many active corridors, but flows are heavily concentrated: a small number of dominant corridors account for the vast majority of migration, while the remainder carry considerably fewer migrants. In period $t + 1$, the system has fewer active corridors, but flows are spread much more evenly: no single corridor dominates. In which period is the system more diverse? The answer depends entirely on which dimension is privileged. If diversity means connectivity, the system is more diverse in period t . If diversity means evenness, the system is more diverse in period $t + 1$. The two periods are genuinely different in kind, and no single number can simultaneously capture both features. This is not merely a technical inconvenience: it reflects a substantive tension in our understanding of migration diversity. The challenge is not

to resolve this tension by selecting one ‘correct’ index, but to adopt a framework that makes the tension visible and tractable.

3.2.3 Hill Diversity Profiles as a Unifying Framework

The diversity measurement problem is not unique to migration. In ecology, the study of species diversity has confronted it most explicitly. An ecosystem with many different species may nonetheless be dominated by one or two abundant species, while most remain rare. For decades, ecologists debated which diversity index was more appropriate, often producing conflicting diversity rankings of the same ecosystems (Daly et al., 2018).

This ecological impasse was resolved by the introduction of Hill diversity profiles (Hill, 1973). Rather than selecting a single measure, Hill’s approach recognised that diversity indices differ primarily in how much weight they assign to rare *vs* dominant categories. By parameterising the weighting through a single sensitivity metric, Hill showed that existing diversity indices are all special cases of a single family of measures: not competing alternatives, but points on a continuous spectrum. At one end of the spectrum, all categories contribute to diversity equally, regardless of size; at the other end, diversity is dominated by the most abundant categories and rare ones contribute negligibly. The resulting ‘diversity profile’—a curve traced out as sensitivity varies—provides a complete picture of distributional shape that no single index can match (Chao et al., 2010; Daly et al., 2018; Jost, 2006; Leinster, 2021). Moreover, Hill profiles can be decomposed into within- and between-category diversity, indicating whether diversity at the aggregate level is the product of individual subpopulations becoming internally more diverse, or more different to each other (Jost, 2007).

The key insight from this literature is that single-number indices conceal weighting choices, while Hill profiles expose them continuously. This allows researchers to see whether conclusions about diversity trends are robust across the full distributional spectrum or specific to particular weighting schemes. If different indices produce

conflicting rankings, as in the previous example with periods t and $t + 1$, the Hill profiles cross and reveal where exactly in the distribution the conflict arises. Neither ranking is wrong; they simply focus on different parts of the distribution and reveal the empirical signatures of distinct structural processes. Disentangling these processes and their sources is the central analytical goal of this chapter.

3.3 Research Questions and Analytical Strategy

By reframing migration diversity as a property of the full corridor distribution, I pursue three related lines of inquiry, each addressing a different level of analysis.

The first concerns the trajectory of global migration diversity since 1960. Previous studies have examined connectivity and evenness separately and reached divergent conclusions. By tracing the full diversity profile over time, it is possible to establish not only whether the global migration system has become more or less diverse, but also at which parts of the distribution diversity has increased or decreased and whether different indices would lead to the same substantive conclusion. This matters because whether the world has become more ‘migratory’ in a meaningful sense depends critically on which part of the distribution one interrogates. The first research question is therefore: (1) *How has global migration diversity evolved since 1960, and do connectivity and evenness tell a consistent story?*

The second line of inquiry concerns the processes underlying global diversity change. Under the distributional perspective, global migration diversity can change through two distinct processes: *corridor turnover*, which refers to the entry (expansion) and exit (contraction) of corridors in the system, and *flow redistribution*, which refers to the reallocation of migration mass across existing corridors. The two processes are conceptually distinct and can in principle operate independently: if new corridors enter the system carrying small flows, they may have little immediate effect on the dominant corridor structure; conversely, a major reallocation of flows can shift the distribution

substantially without activating any new connections. It is also possible that the two processes are mechanically linked: as corridors exit the system, the migration mass they carried may redirect toward remaining corridors, affecting evenness as a by-product of declining connectivity. Whether the two processes operate in tandem or at different times, whether one is mechanically dependent on the other, and which has been the primary driver of global diversity change, are empirical questions the literature has not directly addressed. This gives rise to the second research question: (2) *Was global diversity change driven primarily by corridor turnover or flow redistribution, and did these processes operate simultaneously or at different points in time?*

The third line of inquiry extends the diversity analysis to the country level. Because global diversity aggregates the breadth of individual countries' inflow and outflow portfolios, the same global trend can arise through two fundamentally different country-level dynamics: changes in the diversity of individual countries' portfolios (within-country diversity) and changes in how different these portfolios are from one another (between-country diversity). If global diversity rises because individual countries each develop more balanced migration portfolios, that implies *within-country diversification*: country-level migration systems draw from a wider and more evenly distributed set of partner countries. If it rises because countries increasingly diverge in their partner mix—each concentrating on different origins or destinations—that implies *between-country differentiation*: countries develop distinct migration configurations even as the global system becomes more evenly distributed at the aggregate level.

The two dynamics carry different implications for whether national migration systems are converging or diverging in their networks of linked origins and destinations. Convergence would occur if countries' portfolios became more similar to one another over time: for instance, if all countries drew immigration from, or sent emigration to, an increasingly overlapping set of partners. Divergence would occur if countries became more distinct, each entrenching a characteristic set of dominant partners. The third research question is therefore: (3) *Did changing global diversity reflect within-country*

diversification or between-country differentiation, and what does this imply for convergence or divergence among national migration systems?

Together, the three questions probe whether the apparent paradox in the migration globalisation literature reflects a genuine structural feature of the global migration system or an artefact of how diversity has been measured.

3.4 Data and Methods

3.4.1 Data

To address the three research questions, I construct a dataset of bilateral migration flows covering 195 countries over six ten-year periods from 1960–1970 to 2010–2020. The full list of countries is included in Appendix A.2, Table A.1. Flows are estimated from country-to-country migrant stock data released by the World Bank (2023; Özden et al., 2011), following an established family of indirect estimation methods based on demographic accounting (Abel, 2018; Abel & Cohen, 2019, 2022; Abel & Sander, 2014; Azose & Raftery, 2019).³ The World Bank stock data are measured at mid-year and track international migrants by country of birth, or by country of citizenship where birth information is unavailable (World Bank, 2023).

Crucially for this study’s purposes, the World Bank stock data superimpose current geographies and geopolitical borders onto the historical record: migrants born in one Soviet Republic but resident in another, for instance, are treated as foreign-born prior

³I use flow estimates rather than migrant stocks as the primary data source for two interconnected reasons. First, stocks are cumulative: they reflect the accumulated presence of migrants irrespective of when they arrived, and are therefore slow to register shifts in the corridor distribution. A corridor that was dominant decades ago continues to inflate stock-based diversity measures long after it has ceased to carry large movements. Flows, by contrast, measure the actual reallocation of migrants across corridors within each period and are therefore more responsive to genuine structural change (Ha et al., 2025). Second, stocks conflate distinct types of movement: a person who moves from country A to country B and subsequently returns to A would leave no trace in stock data, yet has completed two migration journeys that are directly relevant to the distributional structure of migration. For these reasons, flows are the more appropriate unit of analysis for this study.

to the dissolution of the Union of Soviet Socialist Republics (USSR; Yildiz & Abel, 2021). This design choice is analytically advantageous for this study. Because the diversity framework proposed here is sensitive to the number of active corridors, any change in political borders that alters corridor definitions would introduce artificial discontinuities into the diversity profile. By anchoring all periods to a consistent geography, the World Bank database ensures that any detected changes in connectivity and evenness reflect genuine structural changes in migration rather than artefacts of shifting boundaries.⁴

Flows are estimated using three alternative demographic accounting methods: minimisation–open, minimisation–closed, and pseudo-Bayesian. I consider as existing flows (and therefore, active corridors) all cases where the estimation resulted in at least one individual moving in a bilateral corridor. The distribution of corridor flows, choropleth maps of country-level total inflows and outflows, and chord diagrams of directional flows (grouped by world region) are reported for each method in the Appendix (see Figures A.1, A.21, and A.30 for the corridor distribution; Figures A.2, A.22, and A.31 for the choropleth maps; and Figures A.3, A.23, and A.32 for the chord diagrams). The main analysis uses the pseudo-Bayesian estimation variant, which best reconciles stock differences with demographic consistency (Abel & Cohen, 2019; Azose & Raftery, 2019). Robustness checks of all empirical results are conducted using the two alternative estimation strategies, as reported in Appendix A.8 and A.9 (Figures A.21 to A.38). Importantly, because the focus is on the properties of the corridor distribution, rather than on absolute flow volumes, differences in volume stemming from alternative estimation methods would not affect the conclusions unless they also produce systematic differences in the shape of the distribution.

⁴This design choice, however, also means that pre-1991 movements between Soviet Republics are counted as international migration, which may inflate connectivity estimates in earlier periods.

3.4.2 Global Hill Diversity Profiles

To address the first research question, I consider a migration system with N active corridors and corridor-level flow shares p_i summing to one. The Hill diversity of order q is then defined as:

$${}^qD = \left(\sum_{i=1}^N p_i^q \right)^{1/(1-q)} \quad \text{for } q \neq 1, \quad (3.1)$$

with the limiting case at $q = 1$:

$${}^1D = \exp \left(- \sum_i p_i \log p_i \right). \quad (3.2)$$

The parameter q governs the sensitivity of the metric to dominant corridors. At $q = 0$, every corridor with positive flow contributes equally, so 0D reduces to connectivity, or the count of active corridors. At $q = 1$, the measure is the exponential of Shannon entropy. At $q = 2$, it is analogous to the migration spread index of Czaika and de Haas (2014), giving disproportionate weight to the largest corridors. As $q \rightarrow \infty$, qD converges to $1/\max_i(p_i)$, depending entirely on the single largest corridor in the system. Importantly, sensitivity to dominant corridors is not a property uniquely located at any single value of q : it increases gradually as q rises, with each value illuminating a different aspect of how migration is spatially organised across dominant and rare corridors.⁵

All Hill numbers are expressed as *effective number of corridors*: ${}^qD = k$ means that the observed distribution has the same diversity at order q as a hypothetical migration system in which exactly k corridors each carry an equal share of migration. This

⁵The parameter q can also take negative values, in which case rare corridors receive greater weight. As $q \rightarrow -\infty$, qD converges to $1/\min(p_i)$, the reciprocal of the relative size of the least popular corridor in the system. In ecology, negative values of q are used to rank ecosystems by the presence of rare or endangered species. In the migration context, however, this specification is less relevant: there is no analogous normative interest in the rarest migration corridors, and the analytical focus throughout this thesis is on how migration mass is distributed across dominant and mid-sized corridors rather than on the long tail of minimal flows.

common unit makes values at different orders directly comparable and gives the measures an intuitive interpretation independent of the migration system's size. Hill numbers also satisfy the replication principle: given two equally large and completely distinct systems, each with diversity D , the diversity of the pooled system is $2D$. Further, because Hill numbers depend only on the relative shares of corridors rather than on absolute flow volumes, they are scale-invariant. For this reason, they are unaffected by changes in overall migration levels and well-suited to period-to-period comparisons.

In addition to the corridor-based measure in Equations (3.1) and (3.2), global diversity can be studied through the marginal distributions of inflows and outflows. By aggregating bilateral flows across origins or destinations, it is possible to obtain measures of inflow and outflow diversity, equivalent, at $q = 2$, to the measures of global immigration and emigration spread in Czaika and de Haas (2014). While corridor diversity captures how migration is distributed across origin–destination pairs, inflow and outflow diversity focus on the composition of migration partners and capture how migration is distributed across destinations or origins.

3.4.3 Country-Level Hill Profiles

All global Hill diversity measures can be implemented at the country level by substituting the corridor-level flow shares p_i in Equations (3.1) and (3.2) with the share of a country's total emigration or immigration accounted for by each partner the country is connected to. At $q = 0$, this gives the count of partners in a country's immigration or emigration portfolio; at $q = 2$, it is equivalent to the country-level emigration and immigration spread indices of Czaika and de Haas (2014). In other words, country-level profiles tell us how diverse a given country's portfolio of connected origins and destinations is. Because all Hill profiles are expressed in the same unit, country-level profiles are directly comparable to one another and to the global profiles.

3.4.4 Top- k Dominance

To complement the global Hill analysis, I compute top- k dominance curves of flow volumes by migration corridors. For ordered flow shares $p_{(1)} \geq p_{(2)} \geq \dots \geq p_{(N)}$:

$$C(k) = \sum_{j=1}^k p_{(j)}. \quad (3.3)$$

This gives the share of global migration accounted for by the largest k corridors. Unlike Hill profiles, dominance curves do not require a sensitivity parameter and make no assumptions about how corridors should be weighted. They are a descriptive complement to the profile analysis: where profiles reveal how diversity changes across the full weighting spectrum, dominance curves show directly which corridors account for the bulk of migration. A steeply rising curve indicates high concentration; a gradual rise indicates more even distribution. Comparing curves across periods reveals whether concentration has shifted toward or away from dominant corridors.

3.4.5 Disentangling Corridor Turnover and Flow Redistribution

To address the second research question, I distinguish between changes in global Hill diversity due to corridors entering and exiting the system (corridor turnover) and to the reallocation of migration mass across active corridors (flow redistribution). I conduct a counterfactual analysis where the contribution of flow redistribution to global diversity is isolated: this is done by restricting pairs of adjacent periods in the sample to the *stable support*, defined as the set of corridors active in both periods. Any remaining difference between total observed diversity change and change on the stable support is attributable to corridor turnover.

Formally, let $\Delta^q D_{t+1} = {}^q D_{t+1} - {}^q D_t$ denote the observed change in diversity between periods t and $t + 1$. For each pair of adjacent periods, I define the stable support as the intersection $S_t \cap S_{t+1}$, that is, the set of corridors active in both periods. Let ${}^q D_t^\cap$

and ${}^qD_{t+1}^\cap$ denote Hill diversity in periods t and $t + 1$ after restricting flows to this stable support and renormalising them to sum to one. Then:

$$\Delta {}^qD_{t+1} = \underbrace{({}^qD_{t+1}^\cap - {}^qD_t^\cap)}_{\text{flow redistribution effect}} + \underbrace{[({}^qD_{t+1} - {}^qD_t) - ({}^qD_{t+1}^\cap - {}^qD_t^\cap)]}_{\text{corridor turnover effect}}. \quad (3.4)$$

The first term captures the contribution of flow redistribution among corridors active in both periods. The second term captures the contribution of corridor turnover, that is, the entry and exit of corridors and the compositional effects these produce. By construction, the two terms sum exactly to the observed diversity change for every value of q .⁶

3.4.6 Within- and Between-Country Diversity Decomposition

To address the third research question and study whether countries' migration portfolios are becoming more internally diverse (within-country diversification) or less similar to one another (between-country differentiation), I examine the diversity of migration inflows and outflows and partition it into within-country (α) and between-country (β) components, following Jost (2007).

For outflows, I consider each origin country c as a distinct subpopulation and define diversity over the distribution of destination countries. Let p_{cd} denote the share of global migration from origin c to destination d , and let $w_c = \sum_d p_{cd}$ denote the share of global migration originating from country c . The within-country component is defined as the weighted power mean of country-level diversities, using order- q weights:

$${}^q\alpha = \left(\frac{\sum_{c=1}^C w_c^q ({}^qD_c)^{1-q}}{\sum_{c=1}^C w_c^q} \right)^{1/(1-q)}, \quad q \neq 1, \quad (3.5)$$

with the limiting case at $q = 1$:

⁶At $q = 0$, Hill diversity reduces to the number of active corridors. Because the stable support contains exactly the same set of corridors in both periods, ${}^0D_{t+1}^\cap = {}^0D_t^\cap$ and the first term is zero. All change at this order is therefore attributable to corridor turnover.

$${}^1\alpha = \exp \left(\sum_{c=1}^C w_c \log {}^1D_c \right). \quad (3.6)$$

The corresponding global diversity is defined over the pooled distribution of destination countries:

$${}^qD = \left(\sum_d \left(\sum_c p_{cd} \right)^q \right)^{1/(1-q)}, \quad (3.7)$$

that is, as the diversity of destinations when all migration flows are aggregated across origins. Between-country diversity is then defined multiplicatively as ${}^q\beta = {}^qD/{}^q\alpha$, yielding ${}^qD = {}^q\alpha \times {}^q\beta$.

The interpretation of each component varies with q . At $q = 0$, α measures the average number of destinations a country connects to, while β captures how unevenly these connections are distributed across countries, reflecting differences in the number of partners that countries have. At higher values of q , α captures how evenly a country's flows are distributed across its major destinations, while β captures whether countries concentrate their flows in the same dominant destinations or in distinct ones.⁷ An increase in α with stable β indicates within-country diversification; an increase in β with stable α indicates between-country differentiation. The same decomposition applies symmetrically to inflows, defining diversity by destination and over the pooled distribution of origin countries.

3.5 Results

To analyse the changing diversity of the global migration system, I proceed in four steps. I first examine how global Hill diversity evolved between 1960 and 2020, showing

⁷At $q = 0$, 0D equals the number of countries that send or receive at least one migrant, and ${}^0\alpha$ equals the average number of partners per country. Because each corridor connects exactly one origin to one destination, the total number of links is the same when counted from the origin or destination perspective. If the set of active origin countries coincides with the set of active destination countries—as is the case in the analysed data—then ${}^0\alpha$, ${}^0\beta$, and 0D are identical for inflows and outflows.

that connectivity and evenness followed divergent trajectories. I then separate the effect of corridor turnover and flow redistribution on diversity change, establishing that the two processes operated at different times. Third, I partition global diversity into within- and between-country components to identify the country-level sources of diversity change. Finally, I trace country-level trajectories for selected cases to illustrate the heterogeneity underlying global patterns.

3.5.1 Two Eras of Global Migration Diversity

I address the first research question by investigating global Hill diversity across all migration corridors in the system, as shown in Figure 3.1. Panel A reports Hill diversity over time, with separate subpanels for different values of the sensitivity parameter q .

At $q = 0$, which measures connectivity, global diversity follows an inverted U-shape: the number of active corridors rose from roughly 19,000 in the 1960s to approximately 22,000 in the 1980s, and then declined steadily to lower levels. At $q \geq 1$, where evenness is emphasised as dominant corridors receive greater weight, the pattern changes markedly. Diversity follows a J-shape: declining slightly through the 1970s, then rising through the 1990s and 2000s to reach its highest levels in the most recent period (2010–2020). At $q = 2$, equivalent to migration spread, the effective number of corridors almost doubled over the six decades, from approximately 160 to 300, meaning that in the 1960s, the global migration system was as diverse as a hypothetical system with 160 equally sized corridors, and by the 2010s, as one with 300.

The full Hill profiles in Panel B corroborate the divergence between connectivity and evenness. The period-specific curves, shown in different colours and changing as a function of q , overlap and cross: this is the empirical signature of structural change operating differently at different segments of the distribution. When all corridors are weighted equally, the 1980s appear most diverse; when dominant corridors receive greater weight, the 2010s appear most diverse.

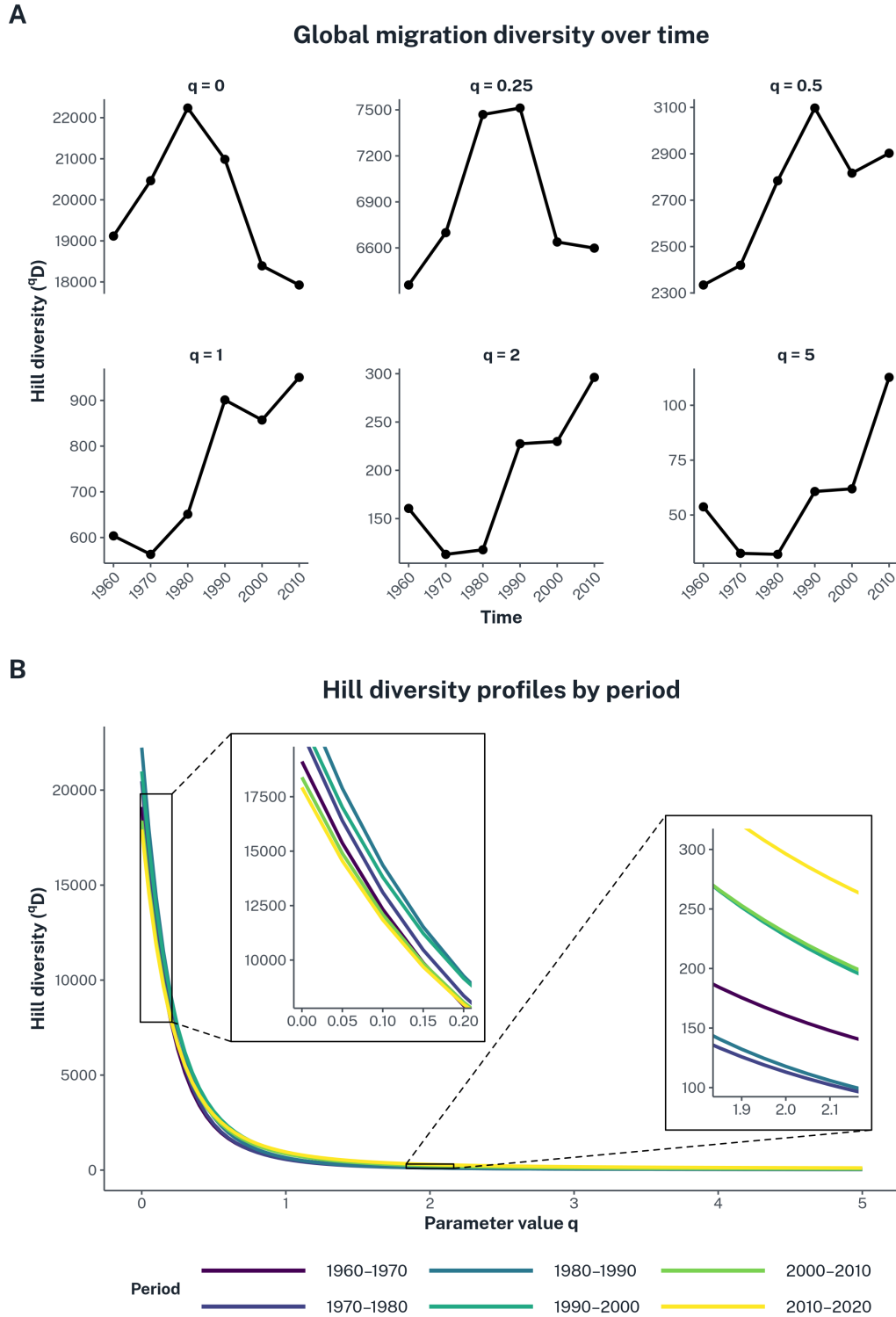


Figure 3.1: Hill diversity of the global migration system, 1960–2020. Panel A shows global Hill diversity over time, indexed by different values of the sensitivity parameter q . At $q = 0$, diversity counts active corridors; at higher values of q , dominant corridors receive greater weight. Panel B shows Hill diversity as a function of q , with different periods represented by differently coloured lines, and highlights the crossing of the curves around $q = 0$ and $q = 2$. In both panels, Hill diversity values are expressed as the number of effective corridors.

The dominance analysis in Figure 3.2 complements this view. Panel A reports the share of total migration accounted for by the ten largest corridors in the system, ranked by flow volume over the whole period. Between 1960 and 1980, the ten largest corridors accounted for 16–20% of all international migration; by 2010–2020, this share had more than halved. The Mexico–US corridor, the largest throughout the period, peaked at nearly 7% of total migration in the 1980s before falling below 2% by 2010. Other major corridors, such as Bangladesh–India and India–Pakistan, similarly lost prominence in more recent periods.

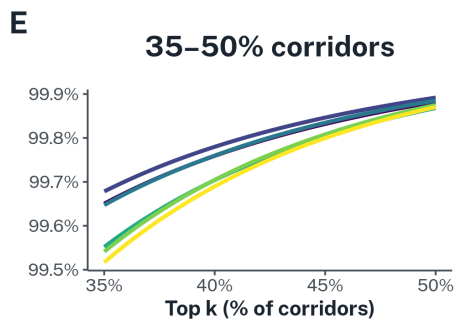
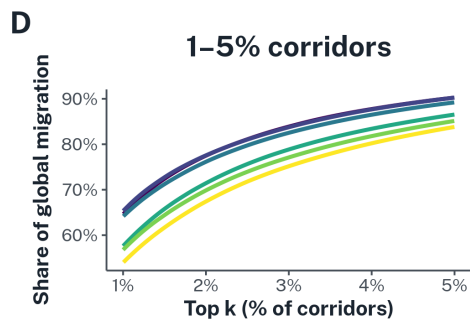
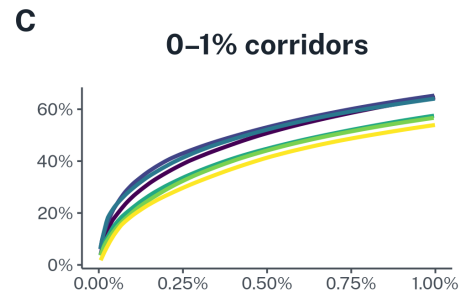
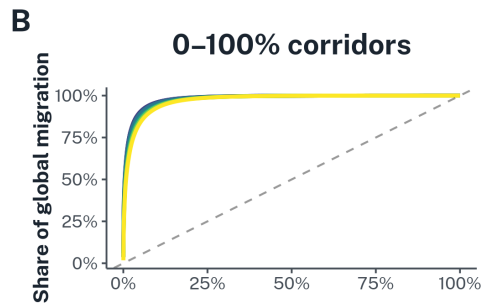
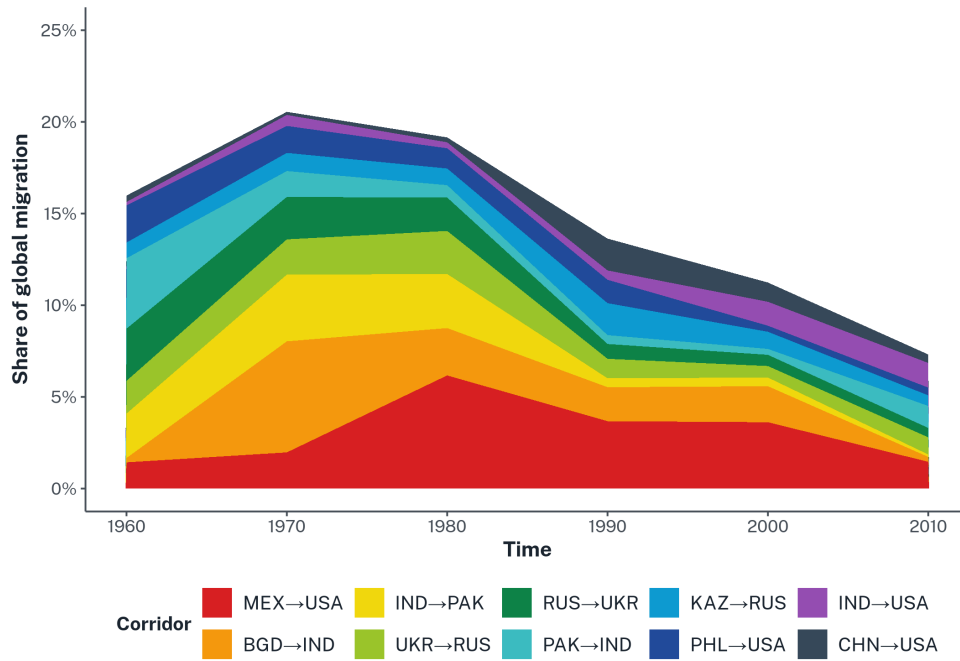
Panels B through E extend this analysis to different segments of the corridor distribution. Panel C shows that the top 1% of corridors accounted for roughly 65% of global migration in earlier periods and less than 50% in recent ones. Panel E reveals a slight reversal for flows in the middle of the distribution (around the 50th percentile): medium-sized corridors carry a somewhat larger share of migration in the 2010s than in the previous decades. This confirms an overall picture of redistribution: migration mass has progressively shifted away from the largest corridors toward smaller and medium-sized ones.

Looking at long-term temporal patterns, Figures 3.1 and 3.2 point to two distinct eras of migration diversity: an era of expanding corridor connectivity through the 1980s, followed by an era of growing evenness from the 1990s onwards, with a structural, progressive weakening of dominant corridors.

3.5.2 Corridor Turnover *vs* Flow Redistribution

A possible reason for the diverging trends in connectivity and evenness is the mechanical reallocation of flows following corridor closure: as corridors exit the system (lower connectivity), flows may redirect towards existing corridors, thereby affecting evenness. To investigate this, I conduct a counterfactual analysis to decompose observed diversity change into a corridor turnover effect and a flow redistribution effect.

A Share of global migration in the top 10 corridors (ranked by total flows)



Period

- 1960–1970
- 1970–1980
- 1980–1990
- 1990–2000
- 2000–2010
- 2010–2020

Figure 3.2: Dominance evolution of corridors in the global migration system, 1960–2020. Panel A shows the share of total migration accounted for by the ten largest corridors over time; panels B–E show the share of total migration accounted for by corridors at different points of the distribution (top 0–100%, 0–1%, 1–5%, and 35–50% of corridors), with different periods reported as differently coloured lines.

The results of this analysis are reported in Figure 3.3, which shows, for selected values of q , the observed changes in global Hill diversity (black solid lines), the changes attributable to corridor turnover (orange dotted lines), and those attributable to flow redistribution (light blue dashed lines).

By construction, at $q = 0$ corridor turnover accounts for all observed change: the rise in diversity through 1980–1990 reflects new corridors entering the system; subsequent decreases reflect net corridor exit. As q increases, however, the pattern reverses.

By $q \geq 1$, diversity gains are almost entirely attributable to flow redistribution. At $q = 2$, corridor turnover contributes nearly zero effective corridors per decade, compared to 50–100 from redistribution. This means that the corridors that entered and exited the global migration system between 1960 and 2010 carried comparatively small flows and had little effect on evenness. Moreover, isolating the effect of flow redistribution reveals that the two processes are empirically decoupled. The growing evenness observed in recent decades is therefore not a consequence of corridor turnover but a distinct structural process operating independently of it. Corridor expansion drove diversity change through the 1980s, while flow redistribution has driven it since the 1990s.

3.5.3 Within- and Between-Country Sources of Diversity

Having established that the post-1990 rise in corridor evenness reflects flow redistribution rather than corridor turnover, I now investigate the country-level sources of changing diversity. I focus on the diversity of origins or destinations when flows are aggregated across all countries (inflow and outflow diversity). This pooled diversity can be decomposed into the average diversity of individual countries' migration portfolios (α) and the degree to which countries differ in their partner mix (β). Figure 3.4 shows this decomposition by inflows (Panel A) and outflows (Panel B) across selected values of q . Within-country diversity trends are shown as solid grey lines, while between-country diversity trends are shown as dashed black lines.

Changes in global migration diversity under observed vs stable corridor set

Diversity change in observed system vs counterfactual with no corridor turnover.

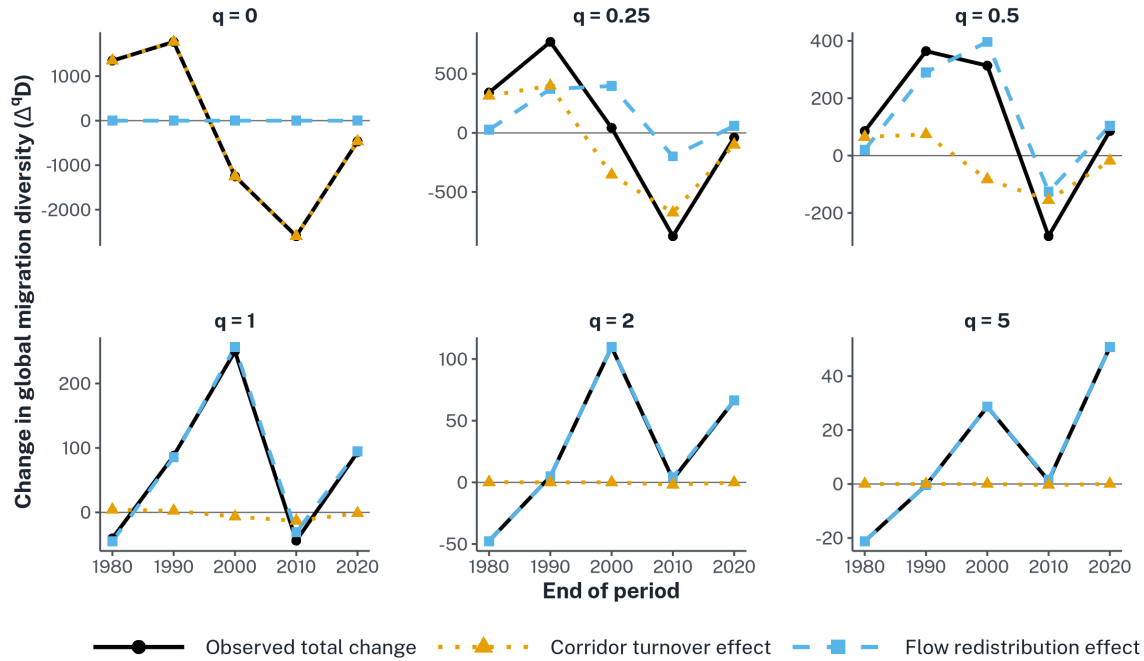


Figure 3.3: Comparison of changes in global migration diversity through a stable-support analysis. Lines show observed changes in global Hill diversity (black solid line); changes attributable to corridors entering and exiting the system (corridor turnover effect; dotted orange line); and changes in the stable set of corridors (flow redistribution effect; dashed light blue line). Hill diversity changes are grouped at selected values of q .

At $q = 0$, inflows and outflows yield identical results by construction, as each connection contributes symmetrically to origin and destination connectivity. Within-country diversity indicates that a typical country connects to approximately 105 migration partners, a figure that remains stable throughout the period. Given that the sample includes 195 countries, this underscores how dense the global migration system is: the average country is connected to over half of all possible partners. Between-country diversity at this order is correspondingly low, remaining below 2 effective sets of non-overlapping partners throughout the period. Because partner sets are so large, they overlap extensively, leaving little room for differentiation between countries at

this order. Similar patterns hold for $0 < q < 1$, where small flows still carry substantial weight and the system continues to exhibit widespread low-volume connections.

At $q \geq 1$, where dominant partners receive greater weight, the picture changes markedly. For inflows, within-country diversity peaks in the 1990s: at $q = 2$, the average destination country receives flows from around 11 effective origins, a sharp reduction from the 105 partners observed at $q = 0$. This shift reflects the concentrated nature of immigration portfolios once small flows are downweighted. At higher orders, within-country diversity declines modestly after the 1990s, before a partial recovery in the most recent period. Between-country inflow diversity follows a different trajectory: it rises through the 1980s, when it exceeds within-country diversity, before declining steadily. At $q = 2$, β reaches approximately 9 distinct portfolios in the 1980s, indicating that global inflow diversity was substantially larger than that of the average country. This shows that the dominant origin partners from which countries receive migrants were highly differentiated across destinations. The peak in the 1980s marks a period of maximum differentiation, after which β declines, indicating that countries began to receive flows from increasingly similar sets of dominant origin partners.

For outflows, within-country diversity follows a U-shaped trajectory, falling to a minimum in the 1970s–1980s before rising to its highest level in 2010–2020. At $q = 2$, the average origin country sends flows to approximately 8 main destinations in the most recent period. Between-country diversity follows the opposite path, peaking in the 1980s and declining thereafter. The gap between α and β is notably larger for outflows than for inflows in earlier periods, indicating that origin countries were more differentiated in their dominant destinations than destination countries were in their dominant origins.

Within- and between-country sources of migration diversity

Partitioning of system-level Hill diversity (qD) into alpha and beta components.

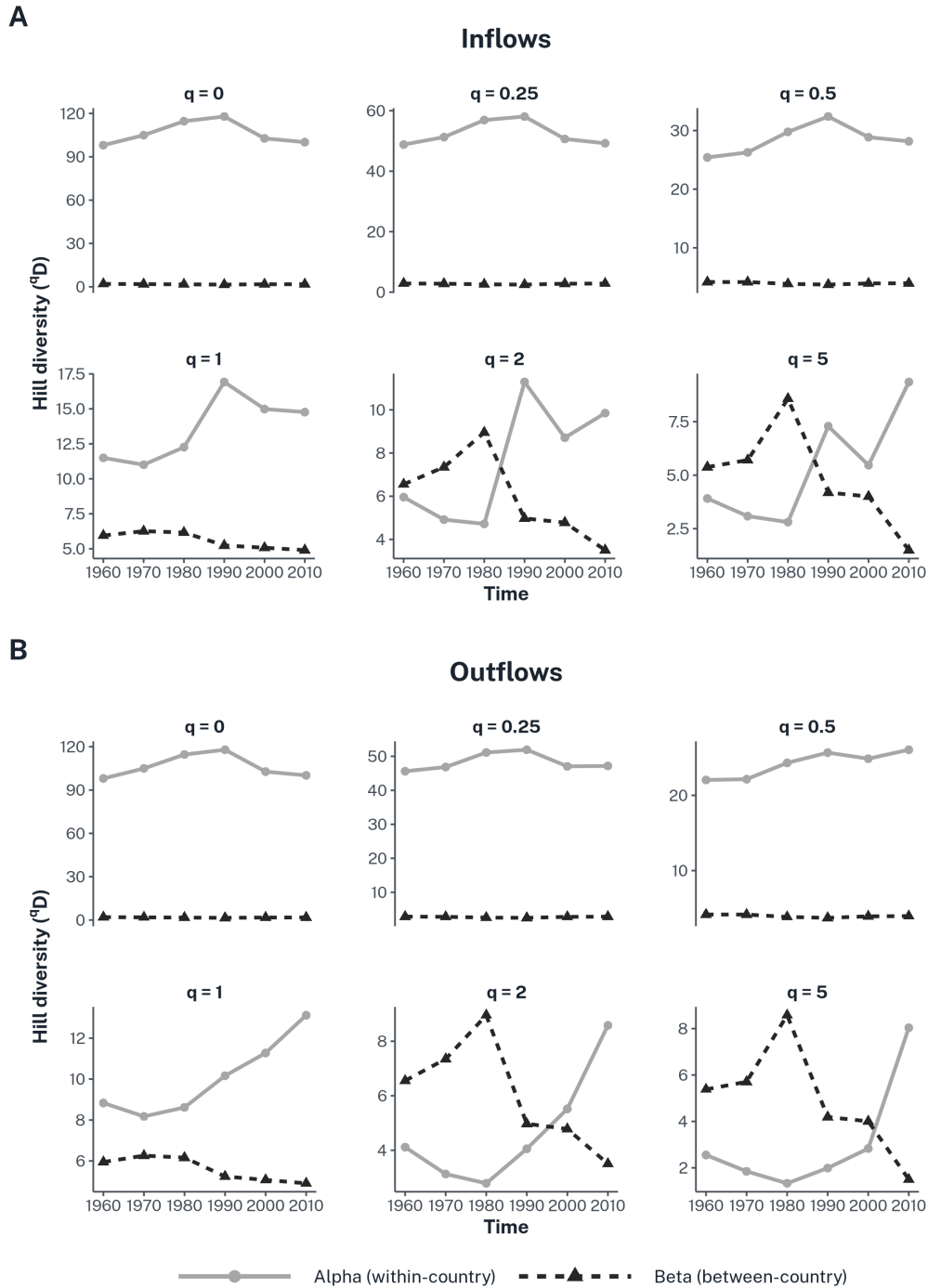


Figure 3.4: Decomposition of global migration diversity into within-country (α) and between-country (β) components. Panel A shows inflow diversity; Panel B outflow diversity. At $q = 0$, α measures the average number of linked partners per country; β how unevenly these links are distributed. At higher q , α reflects how concentrated each country's flows are among its main partners; β how similar or different countries are in their main partners.

This pattern clarifies the country-level sources behind the two eras of global migration diversity. Through the 1980s, between-country diversity rose as countries concentrated on increasingly distinct sets of dominant partners, in a process of between-country differentiation. From the 1990s onwards, this process reversed: between-country diversity declined as countries' migration portfolios became more similar, indicating convergence. Meanwhile, within-country diversity increased, particularly for outflows. Changes in migration diversity since 1990 thus reflect two concurrent dynamics: countries sending and receiving flows more evenly to and from their dominant partners, and increasing overlap in the partners on which they concentrate. This is diversification with convergence: countries have expanded the evenness of their portfolios, while they increasingly share the same dominant partners.

3.5.4 Country-Level Trajectories

The α/β decomposition highlights that diversity changes at the global level are the product of lower-level changes in the networks of origins and destinations in which each country is embedded. A more in-depth analysis of these dynamics requires the inspection of inflow and outflow diversity trends at the country level. I find that country-level diversity trajectories are highly idiosyncratic: as it is not possible to exhaustively investigate the specific trajectories of all countries in the sample, I consider three exemplary countries (Mexico, India, and Italy) and report all remaining country-specific trajectories in Appendix A.6 (Figures A.4 to A.19). The three selected cases illustrate how country-level diversity manifests in highly concentrated sending countries (Mexico), highly diversified sending countries (India), and in countries moving from a primarily sending to a receiving role in the global migration system (Italy).⁸ Figure 3.5 presents the three countries' Hill diversity levels for inflows (solid grey lines) and outflows (dashed black lines) over time and across selected values of q .

⁸The individual contributions of the three countries to global Hill diversity at $q = 0$ and $q = 2$ were further inspected using a leave-one-out strategy, as detailed in Appendix A.7 and reported in Figures A.20, A.29, and A.38.

Diversity of migration inflows and outflows for selected countries

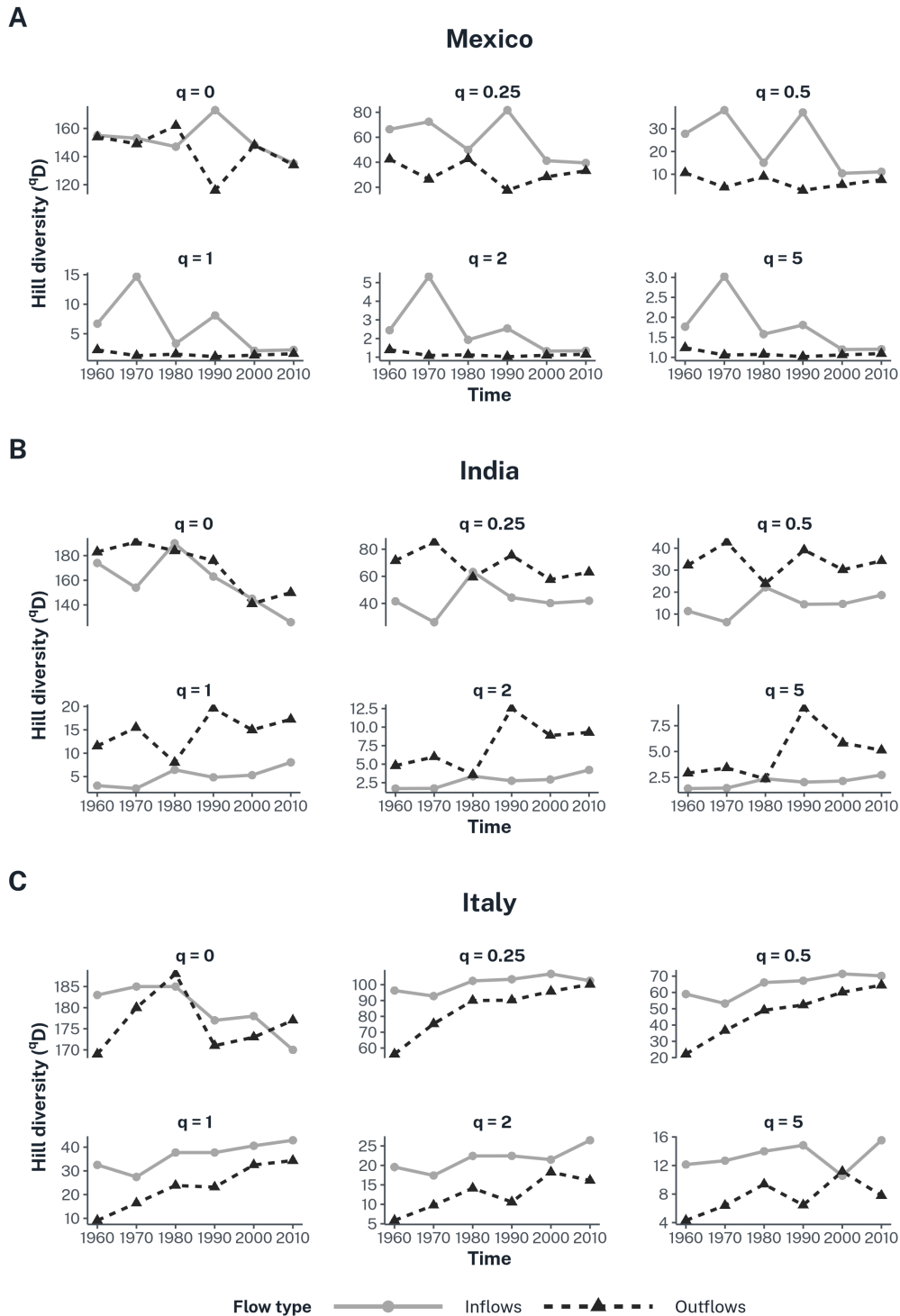


Figure 3.5: Hill diversity of migration inflows and outflows for selected countries, 1960–2020. Solid grey lines show immigration; dashed black lines show emigration. The y-axis shows the effective number of partner countries linked to each country as origins or destinations. Country-level diversity trends are reported for different values of the sensitivity parameter q .

Mexico (Panel A) is a clear example of a country with extreme outflow concentration. At $q = 0$, outflow diversity was relatively stable until 1980, connecting Mexico to around 150 destinations; between 1980 and 1990, this fell below 120, before a partial recovery through corridor expansion in more recent periods. At higher values of q , however, outflow diversity collapses to near one effective destination throughout the study period, with only modest variation. The steep drop as q increases is the signature of an emigration portfolio nominally broad but substantively dominated by a single corridor: in this case, the one to the United States. Mexico's immigration portfolio tells a contrasting story. Inflow diversity is considerably higher than outflow diversity at all $q > 0$, and for much of the study period exceeded it by a factor of three or more. Yet, this gap has narrowed: inflow diversity has declined progressively over time, and by recent periods the two diversities nearly overlap. Mexico's structural position in the global system therefore appears to have shifted: it is increasingly characterised by low inflow diversity alongside persistently concentrated outflows.

India (Panel B) presents a different configuration. At $q = 0$, outflow connectivity was high until more recent periods, connecting India to around 180 destinations in the 1960s before declining to approximately 130 by 2010; inflow connectivity followed a broadly similar trajectory, falling from around 165 to 135 over the same period. At higher values of q , however, the inflow and outflow diversity profiles diverge markedly, and in the opposite direction to Mexico. Outflow diversity rose consistently: at $q = 2$, the effective number of destination countries more than doubled between 1960 and 1990, before stabilising in more recent periods, in line with the growing prominence of Gulf countries as destinations. This underscores the importance of analysing migration diversity beyond volume: despite originating some of the largest migration corridors worldwide, India still succeeded in diversifying its emigration portfolio, progressively redistributing flows across a broader array of destinations. Inflow diversity, by contrast, remains low and flat at all $q > 0$, hovering below 4 effective origins at $q = 2$. This means that India's immigration portfolio has remained

persistently concentrated and dominated by a small number of origin countries, such as Bangladesh, Nepal, and Pakistan.

Italy (Panel C) exemplifies a transition from an emigration to an immigration country. At $q = 0$, inflow connectivity appears stable at around 185 linked origins between 1960 and 1980, while outflow diversity rose rapidly from below 170 destinations in 1960 to over 185 by 1980, briefly converging with the inflow profile. After 1980, both diversities declined: by 1990, outflow connectivity fell to the 1960s levels before recovering slightly in more recent periods; inflow connectivity declined more steadily to around 170 origins by 2010. At higher values of q , the picture is more orderly. Both inflow and outflow diversity grew consistently over the study period, with few oscillations. Throughout, inflow diversity remains consistently higher than outflow diversity: at $q = 2$, Italy is receiving equally sized flows from around 20–25 distinct origins, and sending equally-sized flows to 5–15 destinations. Because both inflow and outflow diversity rise together, Italy exhibits an even redistribution of flows on both the immigration and emigration sides. The earlier crossing of the inflow and outflow profiles at $q = 0$, and their subsequent convergence at higher orders, further mark the structural reconfiguration of Italy's position in the global migration system: from a country whose emigration portfolio was rapidly expanding while immigration was more established, to one where both dimensions are diversifying in parallel.

The three cases also reveal that the within-country diversification of the α/β decomposition emerges from heterogeneous trajectories rather than a uniform process. Mexico's emigration portfolio has remained anchored to a single dominant corridor throughout the period, making it an outlier relative to most other origins and a persistent contributor to between-country differentiation. India illustrates a different pathway: its rising outflow diversity, despite originating some of the world's largest migration corridors, reflects a gradual but asymmetric deconcentration that contributes to the rising within-country diversity observed at the global level. Italy represents a third configuration: at all values of q , Italy's inflow and outflow diversity are

considerably higher than the global α , reflecting a depth of portfolio diversification that extends beyond that of a typical country. Taken together, these cases illustrate how, as with the global-level analysis, a distributional framework of migration diversity makes heterogeneity visible in ways that single indices cannot.

3.6 Discussion

This chapter set out to resolve a persistent ambiguity in the migration globalisation literature: the fact that studies relying on connectivity and those relying on evenness have reached systematically different conclusions about whether international migration has become more or less diverse since 1960. I argued that this inconsistency is not simply a product of data differences or measurement error, but of a deeper distributional problem: any single index of migration diversity takes an implicit position on how much dominant corridors should count relative to peripheral ones, and different positions yield different conclusions. By introducing Hill diversity profiles, I showed that connectivity and evenness are not competing measures but complementary views of the same phenomenon, each illuminating a different part of the spatial organisation of migration flows. Applied to bilateral flow estimates covering 195 countries over six decades, the Hill diversity framework reveals two distinct eras of global migration diversity, identifies the processes underlying each, and locates their country-level sources.

The first research question asked how global migration diversity has evolved since 1960, and whether connectivity and evenness tell a consistent story. The global Hill analysis gives a clear answer: they do not, and this incongruence is not a methodological artefact but a substantive finding about the structure of the global migration system. When all corridors contribute equally to the diversity metric, regardless of size, global diversity follows an inverted U-shape, peaking in the 1980s. When dominant corridors receive greater weight, diversity follows a J-shape, declining slightly through the 1970s

before rising steadily from the 1990s onwards. The crossing of period-specific Hill profiles is the distributional signature of these divergent trajectories: neither the 1980s nor the 2010s can be unambiguously ranked as the more diverse period, because the answer depends on which part of the distribution is privileged. Together, these patterns point to a structural transition in the global migration system, with the 1990s marking a shift from an era of expanding connectivity to one of increasing evenness.

Disagreements in the recent globalisation literature can be understood in these terms. Ha et al. (2025) find no sharp increase in corridor counts between 1990 and 2020, leading them to conclude that migration diversity has plateaued; Fitter et al. (2024) document continued increases in evenness over the same period, suggesting ongoing diversification. Perhaps counterintuitively, both findings are correct. The distributional framework shows that they are not contradictory but complementary: they interrogate different parts of the same distributional object. The literature was not measuring the same thing with different instruments, but different things with instruments that share a label.

The second research question asked whether global diversity change was driven by corridor turnover or flow redistribution, and whether these processes operated simultaneously or at different points in time. The stable-support analysis establishes that the two processes are not only conceptually distinct but empirically decoupled, and that they dominated global diversity change at different historical moments. Through the 1960s and 1970s, corridor expansion was the primary driver: the number of active bilateral connections grew as previously marginal countries became incorporated into the global migration system. From the 1980s onwards, and particularly from the 1990s, flow redistribution became the dominant process: the number of active corridors stabilised and then declined, but flows became more evenly spread across existing connections and dominant corridors lost relative importance. The 1990s thus emerge as a structural pivot: diversity changes in the global migration system went from expansion-driven to redistribution-driven.

The third research question asked whether rising global diversity reflected within-country diversification or between-country differentiation. The α/β decomposition shows that the answer depends on the period under consideration. Until 1990, between-country diversity rose while within-country diversity remained stable: countries concentrated their flows on increasingly distinct sets of dominant partners, and the global system became more diverse through differentiation across national portfolios. From the 1990s onwards, this pattern reversed. Between-country diversity declined as countries started to share dominant partners, while within-country diversity rose. The post-1990 increase in migration diversity thus reflects countries distributing flows more evenly across their main partners, even as those partners increasingly overlap across the global system, in a pattern of within-country diversification and convergence.

This finding intersects with longstanding debates about path dependence in migration systems. Migration has been widely theorised to be self-reinforcing: initial flows create social networks that reduce the costs and risks of subsequent moves, channelling later migrants along established routes (Bakewell, Engbersen et al., 2016; de Haas, 2010; Garip & Asad, 2016; MacDonald & MacDonald, 1964; Massey, 1990). If such feedback mechanisms dominate, one would expect corridor structures to become progressively more entrenched over time, with flows concentrating in established connections rather than spreading across the entire distribution. The dominance curves and Hill profiles show a different trend at the global level: migration mass has moved away from dominant corridors toward smaller and medium-sized ones. The α/β decomposition clarifies how this is possible. In the earlier phase, path-dependent processes contributed to differentiation, as countries became entrenched in distinct sets of migration relationships. In the later phase, these processes coexist with rising within-country evenness and greater overlap in dominant partners across countries. Path dependence may still operate at the corridor level, but its system-wide expression has shifted from between-country differentiation to convergence.

This interpretation also reframes the paradox identified by Czaika and de Haas (2014), where global migration diversity increased between 1960 and 2000 while country-level emigration portfolios became more concentrated. The decomposition suggests that this pattern reflected a particular historical configuration rather than a general property of migration systems. In earlier decades, rising global diversity coincided with between-country differentiation. In more recent decades, global diversity has continued to rise, but now through within-country diversification and declining between-country differentiation. What appeared as a paradox was in fact a transitional configuration: an earlier phase in which countries specialised in different dominant partners, and a later phase in which countries converged on overlapping ones while diversifying internally.

The timing of this transition aligns with major structural transformations in post-World War II migration history. From 1945 to the early 1970s, Western European governments pursued large-scale labour recruitment from the Mediterranean periphery, establishing the 'guestworker' systems that created durable bilateral linkages between origin and destination countries (de Haas et al., 2022). Simultaneously, global-scale decolonisation processes generated new migration corridors: as institutional, linguistic, and socio-cultural ties between former colonies and colonising powers combined with labour demand in destination countries, migration was channelled along post-colonial routes. By the 1990s, which represent an inflection point in this analysis, most post-colonial migration channels had consolidated, and the geographic frontier of corridor formation had largely closed.

A natural question arising from this periodisation is whether the global migration system had, by 1990, approached a ceiling on connectivity: a saturation point beyond which new corridors are unlikely to form in large numbers. The evidence presented here is consistent with such an interpretation: the number of active corridors has declined modestly since its peak, and the counterfactual analysis shows that the diversity gains in recent decades have come almost entirely from flow redistribution

among persistent corridors rather than from corridor turnover. Whether this reflects a durable structural limit or a temporary plateau cannot be determined from the present study alone, but the pattern is striking: the era of corridor expansion appears to have given way to a second era characterised by flow redistribution within an increasingly stable network of migration corridors.

The 1990s pivot also aligns with the collapse of the Soviet Union in 1991, which reshaped the global migration landscape by removing barriers to international movement and establishing new corridors. Crucially, the effect of the Soviet Union collapse on the findings presented here is more subtle than a simple inflation of corridor counts. Because the World Bank data impose contemporary national borders onto the historical record, pre-1991 movements between Soviet Republics are already counted as international migration in earlier periods. What the diversity measure picks up is therefore a distinctive set of new migration configurations among the independent states following the USSR dissolution. Country- and region-specific evidence supports this interpretation, with Russia acting as the primary destination for labour migrants from Central Asia and the Caucasus, Kazakhstan as a regional pole, and the Baltic states reorienting toward Western Europe (Denisenko et al., 2020).

The country cases further underscore the value of studying migration diversity at multiple levels of analysis. Mexico exemplifies extreme and persistent corridor concentration: despite nominal connectivity to many destinations, at higher values of q its emigration portfolio collapses to near one effective destination throughout the study period. This pattern reflects the structural entrenchment of the Mexico–US corridor through half a century of cumulative causation, from the Bracero programme (1942–1964) through subsequent decades of network-driven migration (Garip, 2012, 2019; Garip & Asad, 2016; Kandel & Massey, 2002; Massey, 1990; Massey & Espinosa, 1997; Mata-Codesal & Schmidt, 2020; Palloni et al., 2001; Wassink & Massey, 2022). Hill diversity confirms this, showing that even though Mexico is connected to the world, it substantively sends flows to only one destination.

India presents another configuration that would be invisible without the distributional lens. Like Mexico, India anchors some of the world's largest migration corridors, yet its emigration portfolio has diversified substantially over the study period. This trajectory reflects the emergence of the Persian Gulf as a major destination region from the 1970s onwards. Following the 1973 oil price shocks, labour demand in the Gulf states expanded rapidly, and Indian workers became a major presence across Kuwait, Oman, Qatar, Saudi Arabia, and the United Arab Emirates (UAE; de Haas et al., 2019; Jain, 2016; Khan & Hossain, 2017). The result is rising outflow diversity even as the absolute volume of emigration increased: a pattern that would appear as simple growth in aggregate statistics registers as genuine portfolio diversification in the Hill profiles.

Italy exemplifies a third pathway: a structural role transition from a primarily sending to a receiving country, accompanied by symmetric diversification in both inflows and outflows. For much of the twentieth century, Italy was a major source of emigrants to Northern and Western Europe, the Americas, and Australia. From the late 1970s onwards, this pattern reversed: rapid economic growth, the demographic exhaustion of internal migration from the South of the country, and incorporation into the European free-movement zone transformed Italy into a net immigration destination (Colombo & Dalla-Zuanna, 2019; R. King, 1993). Unlike Northern European countries that relied on bilateral labour agreements with specific origin countries, Italy's 'back door' immigration model—spontaneous matching between labour demand and supply, high rates of irregular entry, periodic regularisation amnesties—drew migrants from a very diverse array of origins: Eastern Europe, North Africa, Latin America, the Philippines, and China (Colombo & Sciortino, 2004). The Hill profiles capture this transformation: when considering the most popular partner countries, inflow and outflow diversity rise consistently and exceed the global α throughout, indicating above-average portfolio breadth on both sides.

The fact that three countries shaped by common global trends exhibit such distinct trajectories underscores a broader point: global aggregates mask substantial heterogeneity at the country level. The mechanisms producing rising evenness operate through the differential entrenchment of national portfolios—each country following its own path-dependent trajectory—rather than through uniform diversification across the system, and country-level trajectories appear highly idiosyncratic. The supplementary analyses in Appendix A.6 document this further. For example, the country-specific diversity trajectories reveal that Syrian emigration became heavily concentrated in the most recent period (2010–2020), which aligns with the outbreak of civil war in 2011. In the same period, Turkey’s inflow diversity decreased considerably, reflecting the sudden rise of forced migration from Syria (de Haas et al., 2022; Seven, 2022). This symmetry in patterns confirms that the proposed diversity framework is diagnostic of structural change in the spatial organisation of flows and highly responsive to external shocks, providing further support for the proposed approach.

Notably, several limitations qualify the present analysis. First, the bilateral flows I rely on are estimates derived from migrant stocks and not register-based counts. This choice was motivated by the fact that no cross-nationally consistent register of migration flows exists that covers all country pairs over the six decades. Validation exercises suggest reasonable alignment between flow estimates and observed data where these are available (Azose & Raftery, 2019), but estimates remain indirect measures that may undercount return, transit, and circular migration. The migration diversity trends presented here are broadly consistent across three estimation strategies, providing reassurance that the main conclusions are not a by-product of the estimation approach, with country-level trajectories showing slightly greater variability. The ten-year aggregation further limits temporal resolution, and the periodisation of the two diversity eras (pre- and post-1990) should thus be interpreted as approximate. The analysis also treats all corridors as equivalent regardless of the type of migration they carry; whether diversification has proceeded in the same way when considering labour

migrants, refugees, students, or family reunification migrants cannot be assessed with the present data.

Most fundamentally, the analysis is descriptive throughout: it establishes what happened to migration diversity and when, but it does not causally identify the mechanisms that produced the observed changes. The decompositions locate change in the distribution and separate its sources; they are not causal models. Candidate mechanisms of diversity change, as discussed above, should be understood as interpretive parallels consistent with the observed patterns and not definitive explanations. At the same time, the framework has diagnostic value and supports evidence triangulation even without causal identification. The Hill profiles and α/β decomposition can detect trend inflections, such as the 1990s pivot, that serve as starting points for further investigation. Identifying *when* and *where* diversity changed is a necessary first step toward understanding *why*, and the distributional framework proposed here can provide a more precise answer to the first two questions.

These limitations point toward productive directions for future research. Extending the Hill diversity framework to incorporate migrant characteristics, such as gender, skill levels, legal status, or duration of stay, would allow researchers to assess whether migration diversification has been spearheaded by particular forms of movement. The finding that diversity trajectories respond sharply to crises suggests that the framework could be applied to study forced displacement specifically, tracking how refugee flows reshape the corridor distribution in real time. Linking the α/β trajectories to corridor-level covariates—shared language, colonial ties, geographic distance, visa agreements—would enable tests of which bilateral characteristics are associated with country-level portfolio differentiation. More ambitiously, combining the distributional framework with causal identification strategies could move the analysis from description to explanation. The Hill profile, the stable-support analysis, and the α/β decomposition are each directly portable to bilateral flow data at any geographic scale, both internal and international, and the decomposition framework can be adapted to partition

diversity by any grouping of interest, making it applicable to questions about regional integration, South–South migration, and the spatial organisation of other forms of human mobility.

More broadly, this chapter underscores the value of treating migration diversity as a distributional property rather than a single number. The history of how diversity has been measured in the globalisation literature has been a history of partial views: corridor counts blind to how flows are distributed across connections, and evenness indices blind to the breadth of the system. The Hill diversity framework does not resolve this by selecting the correct index; it shows that the choice of index is itself informative, because different parts of the distribution have moved in different directions. By making these weighting choices explicit, the framework enables more transparent and comparative assessments of how migration is organised across space and time. If the approach and tools introduced here demonstrate value in this application, they may prove useful wherever researchers seek to characterise the organisation of mobility and understand how system-level patterns emerge from the aggregation of heterogeneous lower-level trajectories.

4 | A Simulation Approach for the Spatial Testing of Migration Theories*

Abstract

Migration research has long been divided between studies of drivers, which focus on the factors shaping migration flows, and studies of patterns, which describe how these flows are organised across space. Theories of migration typically explain migration volumes using drivers, but are less explicit about patterns. As a result, migration theories are usually evaluated using goodness-of-fit measures that assess explanatory power but pay limited attention to spatial accuracy. This chapter addresses this gap by introducing a simulation-based procedure to evaluate the spatial accuracy of migration theories. Starting from an observed system of origin–destination migration flows, the procedure generates synthetic systems that reflect the spatial outcomes implied by a given theory. These synthetic migration systems are then compared to the observed case to assess spatial accuracy. The procedure is applied to intra-European migration flows from 2002 to 2021 and illustrated using two long-standing migration theories: the gravity model and Migration Systems Theory. Both theories achieve high explanatory power under conventional metrics, and migration systems theory performs better overall. However, the empirical analysis shows that both theories fail to reproduce important spatial features of the European migration system, including the high level of reciprocity of flows and the specific migration profiles of Eastern and Northern European countries. These findings highlight how strong statistical fit does not imply accurate spatial representation. Evaluating migration theories through their implied spatial outcomes provides new insights into their limitations and offers a complementary and integrative tool for migration research.

*Material from this chapter has been published as Morellini, M. M. (2026). A Simulation Approach for the Spatial Testing of Migration Theories. *European Journal of Population*, 42(1). DOI: 10.1007/s10680-026-09770-0.

4.1 Introduction

Migration theories carry implicit expectations about how movement should be organised across space: whether flows should concentrate in a few dominant corridors or disperse across many, whether exchanges should be reciprocal or asymmetric. Yet, these spatial implications are rarely tested. The quantitative literature evaluates theories by whether their predictors explain variation in the volume of bilateral flows, not by whether the spatial organisation they imply resembles what is actually observed. The result is a gap between two research traditions that have developed largely in parallel: one focused on explaining why people move, the other on describing how movement is distributed in space.

One fundamental reason for this gap is methodological: while migration drivers are routinely operationalised as predictors in explanatory models, there are few tools for reconnecting drivers to the migration patterns they are presumed to generate. As a result, empirical research has largely assessed theories through conventional goodness-of-fit metrics, which indicate whether predictors help explain volume but provide little insight into their spatial accuracy. Spatial features may be included as control variables, but are seldom treated as substantive outcomes in their own right, leaving unresolved whether competing migration theories provide spatially accurate representations of the migration process.

This chapter offers an alternative approach. By treating migration patterns as emergent outcomes of migration drivers, it evaluates whether theories can produce spatial features consistent with observed data. In doing so, the chapter introduces a methodological framework to evaluate migration theories based not only on the explanatory power and significance of drivers, but also on their spatial accuracy. This is in contrast with standard spatial econometric approaches, where spatial dependencies are treated as explanatory or control inputs; instead, space is regarded as the outcome of interest that theories should be able to reproduce. The proposed method has two main

components. First, it estimates bilateral migration flows using regression models with drivers derived from established migration theories. Second, it simulates counterfactual migration systems under the specified theoretical assumptions and compares simulated and observed spatial configurations, while also quantifying uncertainty.

To demonstrate this approach, I apply it to international migration flows within Europe between 2002 and 2021. Europe provides an ideal empirical setting because it constitutes a relatively controlled institutional environment with extensive, harmonised bilateral migration flow data. The European Union's (EU) freedom of movement regime enables high levels of mobility within its greater borders, creating a quasi-natural experiment in institutionally facilitated international migration. At the same time, substantial cross-country variation in migration volumes has attracted extensive theoretical attention, making Europe a natural laboratory for evaluating how different migration drivers may generate different patterns (Raymer, 2016). In this context, I examine two prominent theories of international migration: the gravity model (Ramos & Suriñach, 2017) and Migration Systems Theory (MST; DeWaard et al., 2012).

The intra-European application shows that, while both the gravity and MST models perform well in terms of conventional explanatory power (eg adjusted R^2), they fall short in reproducing several spatial features of the observed migration system. In particular, the two theories capture the concentration of intra-European migration flows but struggle to reproduce their reciprocity and the specific migration profiles of Eastern and Northern European countries, both in terms of inflows and corridor structures. This means that accounting for volume is not the same as accounting for spatial structure: standard evaluation criteria may validate a theory as empirically supported even when its spatial implications diverge substantially from the observed system.

The chapter's contribution is both conceptual and methodological. Conceptually, it treats spatial patterns as emergent outcomes of theoretical models rather than as nuisance terms to be controlled for, making the spatial implications of migration theories available for direct evaluation. Methodologically, it introduces a simulation-based

procedure—portable across migration contexts and agnostic to specific theories—that generates counterfactual migration systems from fitted models and compares their spatial features to those of the observed system.

The remainder of the chapter reviews the theoretical and methodological background motivating the method, describes the method and its application to intra-European migration, presents the findings, and concludes with a discussion of the results, their limitations, and implications for future research.

4.2 Theoretical and Methodological Background

Once considered demography’s ‘stepchild’ (Kirk, 1960), the study of human migration has evolved into a broad research field covering a large number of subtopics and recognising the contribution of different disciplines, including demography, geography, population ecology, sociology, and economics (Brettell & Hollifield, 2022; Levy et al., 2020; Scholten et al., 2022). Yet, these disciplines have often developed in parallel rather than in dialogue, resulting in what has been described as academic ‘tribalism’ or ‘sectarianism’ (Natter & Welfens, 2024, p. 18).

One consequence of this fragmentation is a persistent divide between research traditions that focus primarily on migration *patterns* and those that focus primarily on migration *drivers*. Work in human geography, spatial demography, and population ecology emphasises migration patterns and the spatial organisation of migration flows (Héran, 2022; M. Price, 2022). By contrast, scholarship in sociology, social demography, and economics predominantly focuses on the drivers of migration, identifying the factors that motivate, enable, or constrain individual and aggregate mobility (Czaika & Reinprecht, 2022; FitzGerald, 2022; Martin, 2022).

Despite being complementary, these two traditions are rarely implemented together in empirical research. Driver-oriented research typically evaluates the significance of predictors stemming from migration theories, but not their spatial accuracy; con-

versely, pattern-oriented research analyses the system-level regularities emerging from migration flows without investigating their social and behavioural origins. I consider each tradition in turn.

4.2.1 Conceptualising Migration Patterns

By connecting physical locations, migration flows serve as spatial linkages. These spatial linkages represent strong empirical regularities: they are the ‘structure’ of migration. Structure can be mapped, measured, and described, usually by examining its *patterns*. Migration patterns thus describe the characteristics of migration structure and, more broadly, of systems (Bakewell, 2024; Boyle et al., 1998; DeWaard & Ha, 2019; Raymer et al., 2019). Migration systems emphasise the spatial organisation of migration and the aggregate (systemic) level of analysis (Bakewell, 2024); as such, they are distinct from migration systems *theory*, which is presented in greater detail in Section 4.2.2.¹

To analyse migration systems and their patterns, researchers have developed a host of system-level descriptive indices. Among these, indices of migration *concentration*, also known as spatial focusing, measure how unevenly migration flows are distributed within a system; in other words, the extent to which migration flows concentrate within a few dominant origin–destination corridors (Plane & Mulligan, 1997; Rogers & Raymer, 1998; Rogers & Sweeney, 1998).² Indices of migration *reciprocity*, also known as migration effectiveness, measure to what extent a migration flow from location i to location j is occurring also in the opposite direction, from j to i (Charles-Edwards et al., 2025; Huang & Butts, 2023; Leal, 2021; Leal & Harder, 2021, 2023; McMillan, 2024).

As empirical regularities and ‘structural imprints’, migration patterns represent the aggregate manifestations of underlying migration processes (DeWaard & Ha, 2019).

¹Migration systems can also be represented by networks, where locations are ‘nodes’ and migration flows the ‘ties’ connecting them (also known as migration ‘macronetworks’; Danchev & Porter, 2021).

²Studies of international migration and the globalisation of migration typically use the complement of concentration indices (1 – index), so that higher values indicate greater evenness or ‘spread’ (Czaika & de Haas, 2014).

This idea forms the basis of the methodology proposed here, which evaluates theories in terms of the spatial patterns they imply.

4.2.2 Conceptualising Migration Drivers

In the international migration literature, drivers refer to the factors that motivate or constrain changes in residence or moves away from one's country of origin (conventionally, country of birth or nationality). Migration drivers include micro-level factors, such as individual aspirations, motivations, and capabilities; meso-level factors, such as interpersonal networks and other intervening factors; and macro-level factors such as geographic, historical, and politico-institutional conditions (Czaika & Reinprecht, 2022; de Haas, 2021).

Traditionally, migration theories are operationalised as sets of migration drivers, which are then used as predictors of migration flows (the outcome variable) in explanatory models (Garip, 2012). One of the classic models exemplifying this approach is the gravity model, first developed by Ravenstein (1885, 1889). The gravity model posits that migration between two locations increases with population size and decreases with the geographic distance between them. Though simple, the gravity model has proven durable and remains widely used across demography, economics, and sociology (see for instance Botezat & Ramos, 2020; Lewer & Van den Berg, 2008; Poot et al., 2016; Ramos, 2016; Ramos & Suriñach, 2017).

A seminal account of how different migration theories can be operationalised with drivers is provided by Massey et al. (1993). Massey et al. survey eight major theories of migration: the neoclassical economic model, the new economics of labour migration, dual labour market theory, world-systems theory, network theory, institutional theory, cumulative causation, and MST. In all cases, the authors 'translate' the abstract notions and mechanisms of theories into measurable and testable drivers. Among the surveyed theories, the authors identify MST as the most comprehensive one, as it synthesises mechanisms from many of the preceding theories and integrates a wide array of

geographic, economic, institutional, historical, and socio-cultural linkages to explain the persistence and spatial organisation of flows.

In practice, MST is operationalised as a wide set of predictors, including: (i) geographic predictors such as distance and contiguity, akin to gravity models; (ii) economic differentials between origins and destinations, drawing from neoclassical and new economics perspectives; (iii) prior migration flows and stocks, consistent with theories of network migration and cumulative causation; and (iv) predictors stemming from world-systems and institutional theories, such as colonial and other historical relationships between locations, shared language, and institutional membership (DeWaard et al., 2012; Leal, 2021; Leal & Harder, 2021, 2023; Massey et al., 1993).

While many migration theories are potentially relevant for empirical testing, this chapter focuses on the gravity model and MST because they represent two ends of a spectrum of theoretical sophistication. The gravity model is parsimonious and centres on a small number of predictors (population size and geographic distance), whereas MST is more elaborate and integrates a broader set of economic, institutional, historical, and relational factors. Notably, both theories assign primary importance to macro-level drivers and operate at the same analytical level, making them directly comparable within a unified modelling framework. At the same time, they differ in the extent to which they articulate expectations about the spatial organisation of migration systems. In the gravity model, migration patterns emerge implicitly from dyadic and largely symmetric attributes: flows are shaped by distance and size, but the theory itself makes few explicit claims about the resulting system-level spatial organisation. Considering its main drivers, we might conclude that gravity would imply relatively more even and reciprocal patterns of exchange than other theories, though not necessarily fully symmetric ones. MST, by contrast, is more explicit in its spatial expectations. It conceptualises migration systems as structured by hierarchical and historically embedded relations, anticipating core–periphery

configurations characterised by concentrated flows, directional asymmetries, and low reciprocity (Leal, 2021; Leal & Harder, 2021, 2023; Windzio, 2018).

The difference in how explicitly spatial structure is theorised makes gravity and MST particularly well suited for comparative evaluation. By assessing not only how well each model explains bilateral flows, but also how accurately it reproduces structure, the analysis moves beyond standard goodness-of-fit metrics and toward a more comprehensive, spatially explicit assessment of theories.

4.2.3 Closing the Gap Between Migration Patterns and Drivers

The contrasting spatial expectations of gravity and MST point to a broader methodological challenge: despite clear theoretical links between migration drivers and patterns, most empirical approaches inspect them separately. While recent scholarship has called for more spatially aware approaches to migration modelling, existing quantitative frameworks still fall short of integrating these dimensions (Czaika et al., 2025; Hoffmann et al., 2021; Matthews et al., 2021; Ton et al., 2024).

Prominent attempts to bring space into empirical migration analysis can be found in the spatial econometrics literature, where models incorporate spatial dependence through spatially lagged predictors or correlated error structures (Elhorst, 2013). While these models account for autocorrelation and improve statistical inference, they conceptualise space primarily as a predictor, correction term, or mechanical dependence structure. Spatial components are introduced to reduce bias or model spillovers, not to analyse migration structure as the outcome of interest. Consequently, the statistical significance of individual predictors tells us little about a model's spatial accuracy: in the migration case, it does not tell us whether a given combination of drivers, even when well-specified and significant, can reproduce the spatial configuration of observed migration systems. For the same reason, existing approaches offer limited tools for assessing uncertainty around spatial accuracy, making it difficult to evaluate spatial robustness (Matthews et al., 2021).

To address these gaps, I propose a simulation-based method that treats migration patterns as emergent outcomes of theoretical models, rather than as nuisance terms or variables to be controlled for. By (i) estimating flows under competing sets of drivers; (ii) simulating full counterfactual migration systems; (iii) comparing their spatial patterns to those of the observed system; and (iv) incorporating uncertainty, the method directly evaluates the spatial accuracy of migration theories. In doing so, it responds to calls for methodologies that ‘bring together spatial thinking and demographic thinking’ (Raymer, in Matthews et al., 2021, p. 5) and provides a unified framework for adjudicating between alternative explanations on both statistical and spatial grounds, thereby allowing for a more comprehensive assessment of which theories best reproduce the architecture of real-world migration systems.

4.2.4 Simulation Approaches in Migration Research

The simulation-based method developed in this chapter builds on advances in computational social science, particularly in simulation modelling. Over the past twenty years, simulations have become increasingly widespread in demography and migration studies (Bilecen et al., 2018; Drouhot et al., 2023; Kashyap & Zagheni, 2023; Kashyap et al., 2023). Simulation approaches are usually divided between micro-level simulations, which include Agent-Based Models (ABMs) and microsimulations, and macro-level simulations, which include system dynamics and complex networks (Kashyap & Villavicencio, 2016; Margetts & Dorobantu, 2023; Raftery, 2000). As the specific differences between micro and macro simulation approaches are beyond the scope of this chapter, the remainder will discuss simulation strategies more broadly.

The relevant contribution of both micro- and macro-level simulations to the proposed method is that they detect theorised processes underlying emergent patterns, based on individual (micro) or system (macro) behaviours (Drouhot et al., 2023; Klabunde & Willekens, 2016). In practice, simulation methods in migration research consist of two core ingredients. First, they define a set of entities—such as individuals,

households, or locations—and of a finite set of states these entities can occupy (eg residence in a given location, migrant *vs* non-migrant status). Second, they specify transition rules to regulate how entities move between states over time. These rules are typically probabilistic and derived from empirical models, stylised behavioural assumptions, or secondary data. Simulation approaches can thus be seen as multilevel and multistate, as agents and systems are allowed to transition between various states (Bijak et al., 2018).

At the micro-level, simulations have been used to explore how environmental change shapes rural–urban and international migration decisions (Kniveton et al., 2011) and how social networks sustain and amplify migration streams over time (Klabunde & Willekens, 2016). At the macro level, simulations have been applied to examine how migration routes emerge and stabilise over time (Bijak, 2022) and how internal mobility flows respond to economic and policy shocks (Huang & Butts, 2024).

Both micro- and macro-level simulations can be used to generate synthetic counterfactual and ‘what if’ scenarios, which can be particularly valuable for investigating causal relationships when experimental designs such as randomised controlled trials may be infeasible or ethically inappropriate (Kashyap & Zagheni, 2023). For the same reason, simulations can be used for scenario exploration and hypothesis testing. While applications of this kind remain somewhat rare in demography and migration studies, a broader literature in sociology motivates their implementation as part of the proposed method.

4.2.5 Simulations for Hypothesis Testing

One of the most important research pieces establishing the connection between simulations and hypothesis testing in the social sciences is by G. King et al. (2000). In their article, the authors introduce simulations as a powerful tool to extract insights from empirical models and make findings accessible to non-technical audiences. The authors posit that social scientists tend to focus too narrowly on technical terms such

as p values and statistical significance, limiting the wider interpretability of findings. By contrast, simulations enable researchers to present results as precise estimates of real-world quantities and counterfactuals. The use of simulations for hypothesis testing thus offers three major benefits. First, it allows researchers to extract theoretically meaningful quantities from statistical models. Second, it clarifies, also visually, the uncertainty surrounding estimates without referring to technical terms. Lastly, it makes complex findings accessible to a broader audience without requiring specialised statistical knowledge.

In sociology, the simulation-based workflow of G. King et al. (2000) has inspired applications across several domains. For example, Alon and Gelbgiser (2011) use post-estimation simulations to quantify how horizontal sex segregation across fields of study contributes to gender gaps in college completion. Block (2023) adapts the simulation framework to social mobility networks to examine how individual transitions in occupations aggregate into gender segregation in the labour market. Similarly, Huang and Butts (2024, 2023) employ macro-simulations to examine how diverging tendencies towards internal migration or immobility can produce spatial and social segregation at the aggregate level of US counties. Across these studies, simulations serve primarily to derive interpretable quantities from statistical models, explore counterfactual scenarios, or examine how assumed behavioural rules translate into macro-level patterns.

The approach proposed here draws on and extends these strategies. Like previous work, it uses simulations to generate counterfactual systems under alternative assumptions. However, rather than simulating behavioural outcomes conditional on a fixed network or population configuration, the method simulates entire bilateral migration systems implied by competing theoretical models, explicitly incorporating simulation-based uncertainty around spatial accuracy. This allows researchers to assess whether the spatial structures produced by a given migration theory align with

those observed in the data, broadening simulation-based inference for the purpose of spatial theory-testing.

4.3 Proposed Methodology: A Spatial Test for Migration Analysis

The method developed here evaluates the spatial accuracy of migration models by linking migration drivers to observed patterns. It consists of four main steps, summarised in Figure 4.1. These steps are designed to be flexible and adaptable, with the methodological contribution lying primarily in the *sequence* of operations, not in the specific modelling choices. This ensures applicability across a wide range of migration contexts and data structures.

To illustrate the approach, I present one possible implementation, which relies on: a regression-based model (Step 1), a computationally efficient simulation technique (Step 2), established spatial indices from migration research (Step 3), and Mahalanobis distance for model evaluation (Step 4). The simulation technique used here is designed for scalability: in the empirical application on intra-European migration, millions of individuals across 31 countries over 20 years are simulated in seconds on a standard workstation using the statistical software R (version 4.3.2; R Core Team, 2023).

4.3.1 Step 1: Model Specification

The first step of the procedure requires estimating a statistical model of migration, in which bilateral flows are explained as a function of theory-derived drivers. Different theoretical frameworks can be operationalised as alternative sets of predictor variables. An explanatory model of migration flows would then take the following functional form:

$$y_{ij} = \beta x_{ij} + \varepsilon_{ij}, \quad (4.1)$$

where the dependent variable y_{ij} is the migration flow from an origin i to a destination j , estimated as a function of predictor variables in the vector x_{ij} , plus an error term, ε_{ij} . The model's coefficients are included in the vector β .

Step 1 translates theoretical assumptions into testable model inputs. Each set of predictors represents not only a claim about what drives migration flows, but also an implicit prediction about what the resulting spatial organisation should look like.³

4.3.2 Step 2: Simulating Migration Systems

Step 2 implements simulation techniques to generate alternative, theory-guided migration scenarios. To ensure direct comparability with the migration drivers specified in Step 1, the simulations are based on the fitted values from the regression model in Equation (4.1). Following the logic of simulation-based inference introduced by G. King et al. (2000), uncertainty around the estimated migration flows is propagated into the spatial domain by repeatedly simulating complete migration systems implied by the fitted model.

In practice, simulations are carried out at the level of bilateral flows. For each origin–destination pair (i, j) and year, a fitted flow $\hat{y}_{ij}$ and its associated confidence interval are obtained from the regression model. Assuming approximate normality of the sampling distribution of the fitted values, the width of the confidence interval is used to recover a standard deviation for each dyad. Simulated flows are then drawn independently for each dyad from a truncated normal distribution:

$$y_{ij}^* \sim \mathcal{TN}(\hat{y}_{ij}, \sigma_{ij}^2; a_{ij}, b_{ij}), \quad (4.2)$$

³In this step, it is recommended to account for data interdependencies that are common in migration research. Estimators such as Poisson Pseudo-Maximum Likelihood (PPML) with origin and destination fixed effects (Santos Silva & Tenreyro, 2006), or network regressions with Quadratic Assignment Procedure (QAP; Barnett & Nam, 2024; Plotnikova & Ulceluse, 2022) can address these dependencies.

where σ_{ij} is derived from the confidence interval bounds (a_{ij}, b_{ij}) . Truncation ensures that simulated flows remain within empirically plausible ranges and prevents the generation of negative migration values.

Each draw across all dyads constitutes a single simulated migration system, represented by an origin–destination flow matrix. To ensure close correspondence with the observed migration system, simulated flows can be calibrated so that total outflows from each origin match the observed outflows in the corresponding year. This row-wise scaling preserves the set of observed origins and destinations, prevents the appearance of migration corridors that are not present in the original data, and ensures that the aggregate volume of migration in each simulated system matches that of the observed system.

Repeating this procedure yields an ensemble of counterfactual migration systems implied by the fitted model. Rather than producing point predictions for individual flows, the simulation strategy generates a distribution of possible migration systems consistent with the estimated drivers and their uncertainty. This ensemble provides a sampling distribution for both flow-level quantities and system-level spatial features, which are used in Step 4 to assess how unusual the observed migration system is relative to what the regression model would imply.

The simulation strategy adopted here is computationally efficient and well suited for testing theories that emphasise macro-level drivers, such as the gravity model and MST. A limitation of this approach is that dyadic flows are simulated independently and do not explicitly model individual destination-choice behaviour. While this makes the approach scalable to large migration systems, it abstracts from micro-level decision-making processes. Nonetheless, Step 2 can be implemented using individual-level simulation strategies as well, including microsimulations and ABMs, allowing for agentic destination selection, albeit at a substantially higher computational cost.

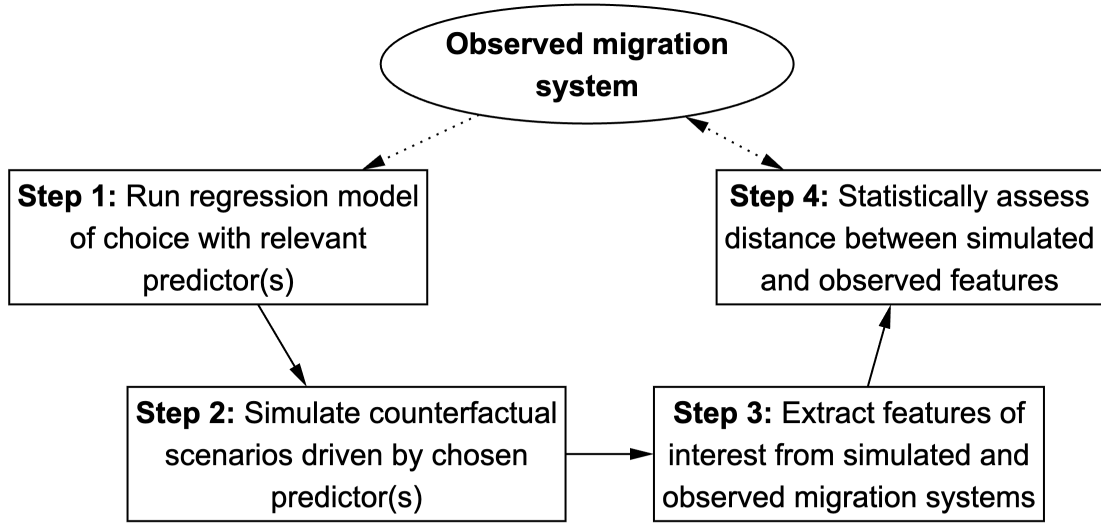


Figure 4.1: Diagram of the proposed simulation-based procedure.

4.3.3 Step 3: Extracting Spatial Features

Step 3 extracts spatial features of interest from the simulated migration systems, enabling direct comparison with observed data. Here, I rely on four migration indices, but alternative or additional indices could be employed to describe migration patterns.

In most cases, and all those considered here, migration indices are computed using matrix notation, where a migration system, M , is a matrix with rows representing the origins and columns the destinations, so that an element m_{ij} will report the total number of people moving from location i to location j . If, between two locations, no migration flow is observed in that specific direction, the corresponding matrix element will be equal to zero. By nature of most migration analyses, where the focus is on individuals who decide to relocate, the diagonal of the origin–destination matrix is usually set equal to zero (Raymer & Rogers, 2007).⁴

⁴In the case of longitudinal data, a migration matrix, its elements, and any descriptive index can refer to a specific time period, t . For clarity and simplicity, this temporal subscript is not reported in the following mathematical notations and formulae.

Measures of Migration Concentration

Indices of migration concentration quantify how unevenly migration flows are distributed across a system: whether flows are spread in similar volumes across many corridors or concentrated within a few dominant ones. Because different indices weight parts of the flow distribution differently, three complementary measures are used here to triangulate migration concentration: the Aggregate Coefficient of Variation (ACV), the Gini Index, and the Migration Inequality Index (Rogers & Raymer, 1998).

Aggregate Coefficient of Variation (ACV) The ACV, first applied to migration by Rogers and Raymer (1998) and Rogers and Sweeney (1998), measures the relative dispersion of flows as the ratio of the standard deviation to the mean. Applied to a migration matrix $\mathbf{M}$ with n locations, it is:

$$\text{ACV} = \sum_{i=1}^n \left(\frac{\sqrt{\frac{1}{n-1} \sum_{j \neq i} (m_{ij} - \bar{m}_i)^2}}{\bar{m}_i} \right) \cdot \frac{\sum_{j \neq i} m_{ij}}{\sum_{i=1}^n \sum_{j \neq i} m_{ij}} \quad (4.3)$$

where m_{ij} is the flow from i to j and $\bar{m}_i$ is the mean outflow from i . The ACV has no upper bound and is highly sensitive to the presence of one or two disproportionately large corridors, also known as primacy. Rogers and Raymer (1998) emphasise that this sensitivity helps distinguish systems where a few flows dominate from those where migration mass is more evenly distributed.

Gini Index The Gini Index of migration measures concentration across all bilateral corridors and ranges from zero (no concentration) to one (maximal concentration; Bell et al., 2002; Plane & Mulligan, 1997):

$$\text{Gini} = \frac{\sum_{i=1}^n \sum_{j \neq i} \sum_{k=1}^n \sum_{l \neq k} |m_{ij} - m_{kl}|}{(2n(n-1) - 1) \sum_{i=1}^n \sum_{j \neq i} m_{ij}} \quad (4.4)$$

where each flow m_{ij} is compared to every other flow m_{kl} in the system. The denominator normalises the index so that it ranges between zero and one. Compared to the ACV, the Gini Index is less influenced by extreme values and more sensitive to changes in the middle of the flow distribution (Rogers & Raymer, 1998, p. 65). This makes it a useful complement for assessing overall inequality without undue weighting toward the largest corridor(s).

Migration Inequality Index The Migration Inequality Index measures the deviation of observed flows from a hypothetical uniform distribution, in which all corridors would contain the same number of migrants (Bell et al., 2002). It ranges from zero (uniform distribution; all migration corridors equally used) to one (maximal concentration; only one corridor used):

$$\text{Inequality} = \frac{\sum_{i=1}^n \sum_{j \neq i} |m_{ij} - m'_{ij}|}{2 \sum_{i=1}^n \sum_{j \neq i} m_{ij}} \quad (4.5)$$

where m'_{ij} represents the expected flow under a uniform distribution. While conceptually similar to the Gini, this index is more strongly affected by outliers and large deviations from the expected flow distribution, and therefore sits between the ACV and the Gini.

Measures of Migration Reciprocity

Indices of migration reciprocity measure whether and to what extent migration flows are symmetric: in other words, whether flows occur in both directions of a migration corridor, from location i to location j and from location j to location i (Huang & Butts, 2023; Leal, 2021; Leal & Harder, 2021, 2023; McMillan, 2024). Following Krivitsky (2012), the reciprocity of a migration system can be computed by comparing the two flows in each corridor and reducing them to a single value using their minimum, maximum, or geometric mean, and then aggregating across all corridors.

Here, I use the minimum specification, which focuses on the smaller of the two directional flows in each corridor:

$$\text{Reciprocity} = \frac{\sum_{i=1}^n \sum_{j \neq i} \min(m_{ij}, m_{ji})}{\sum_{i=1}^n \sum_{j \neq i} m_{ij}}. \quad (4.6)$$

The denominator normalises the measure so that it ranges between zero and one. Substantively, the numerator counts the number of migrants who can be ‘matched’ to someone moving in the opposite direction of the corridor, while the denominator is the total number of migrants in the system. The index can therefore be interpreted as the share of migrants who are part of reciprocal exchanges: for example, a value of 0.7 indicates that 70% of all migrants belong to corridors where flows occur bidirectionally. Values near zero indicate highly asymmetric systems dominated by one-way flows, whereas values near one indicate highly reciprocal systems in which most corridors are approximately balanced.

4.3.4 Step 4: Statistical Assessment of Spatial Fit

The simulated migration systems produced in Step 2 represent theory-derived counterfactuals: they show what the spatial structure of migration would look like if flows were driven solely by the predictors specified in Step 1. Step 4 evaluates the spatial accuracy of these theoretical models by comparing the observed migration system with the ensemble of simulated systems.

Building on the logic of random networks and Conditional Uniform Graph (CUG) tests, Step 4 repeats the sampling process described in Step 2 to generate a reference distribution of spatial features under each theoretical model. The comparison between observed and simulated features can therefore be interpreted as a spatial test of goodness-of-fit. This approach is analogous in spirit to simulation-based assessments of Exponential Random Graph Models (ERGM; Lusher et al., 2012), but allows for more calibrated modelling of bilateral migration flows. Because each spatial characteristic is

analysed across all simulated systems, the method evaluates not a single point estimate but the full distribution of spatial outcomes implied by each theoretical model.

Model evaluation proceeds in two complementary ways. First, visual inspection of the simulated distributions provides an intuitive assessment of how well each model reproduces the observed spatial features; graphical diagnostics are particularly useful for identifying patterns of model misfit (see G. King et al., 2000). Second, statistical testing is carried out using distance measures that quantify discrepancies between observed and simulated systems. Here, I rely on the Mahalanobis distance, which has been used extensively to evaluate the accuracy of statistical network models (Lospinoso & Snijders, 2019). In the migration context, Mahalanobis distance can be applied either to migration indices (eg concentration or reciprocity) or to full vectors of flows for each origin or destination. This allows researchers to identify not only which spatial features are poorly captured by a model, but also which specific locations contribute most to the misfit.

Formally, for each location or migration corridor, we compare an observed flow vector, $\mathbf{x}$, which can represent inflows, outflows, or both, to the distribution of simulated flow vectors ($s = 1, 2, \dots, S$). Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the mean and covariance matrix of the simulated vectors; the Mahalanobis distance is then:

$$D^2 = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}). \quad (4.7)$$

By incorporating the inverse covariance matrix, Mahalanobis distance accounts for dependencies between origin–destination pairs, providing a multivariate measure of how unusual the observed flows are relative to the simulated distribution. Large values of D^2 indicate that the observed flows differ substantially from what the model predicts, whereas values near zero indicate close agreement. In longitudinal analyses, this metric can also be tracked to assess whether spatial accuracy improves or deteriorates over time.

4.4 Application: International Migration in Europe, 2002–2021

To demonstrate how the proposed procedure can be used in practice, I apply it to international migration in 31 European countries from 2002 to 2021.⁵ The European case is particularly well-suited for this purpose for four main reasons.

First, Europe offers a relatively controlled institutional environment with high levels of cross-border mobility. The EU Treaty of Maastricht (1992) and the Free Movement Directive (in force since 2004) grant citizens of EU member states reciprocal rights to move, reside, and work across borders, creating a quasi-natural experiment in free international migration within the EU's wider boundaries (European Commission, 2023; Maciejewski, 2024; Marzocchi, 2024). Further bilateral agreements have been established with non-EU countries in the continent, including the Schengen Agreement (first established in 1985) and the Agreement on the Free Movement of Persons (AFMP) between Switzerland and the EU (2002). This symmetric mobility rights regime is especially relevant for assessing spatial features such as reciprocity.

Second, Europe is an exceptionally data-rich context (Raymer, 2016). Migration inflows and outflows in the region are recorded by national statistical offices as well as supranational institutions such as Eurostat, and Nordic countries in particular provide high-quality register data (Daňko et al., 2024). The availability of multiple data sources facilitates internal consistency checks between information provided by sending and receiving countries, supporting the construction of harmonised migration matrices (Daňko et al., 2024; Kupiszewska & Wiśniowski, 2009).

Third, the period under consideration offers annual flow data over two decades with stable national borders, allowing for consistent longitudinal and cross-country

⁵The selected countries are (in alphabetical order): Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom (UK). See also Appendix B.2, Table B.1.

comparisons of migration patterns. This is in contrast to other migration cases, such as internal migration, where data may be less frequent and geographical units change over time (Lomax, 2022).

Finally, intra-European migration has been a major testing ground for macro-level theories, particularly the gravity model (Ramos & Suriñach, 2017) and MST (DeWaard et al., 2012), which provide the two driver specifications evaluated in this application.

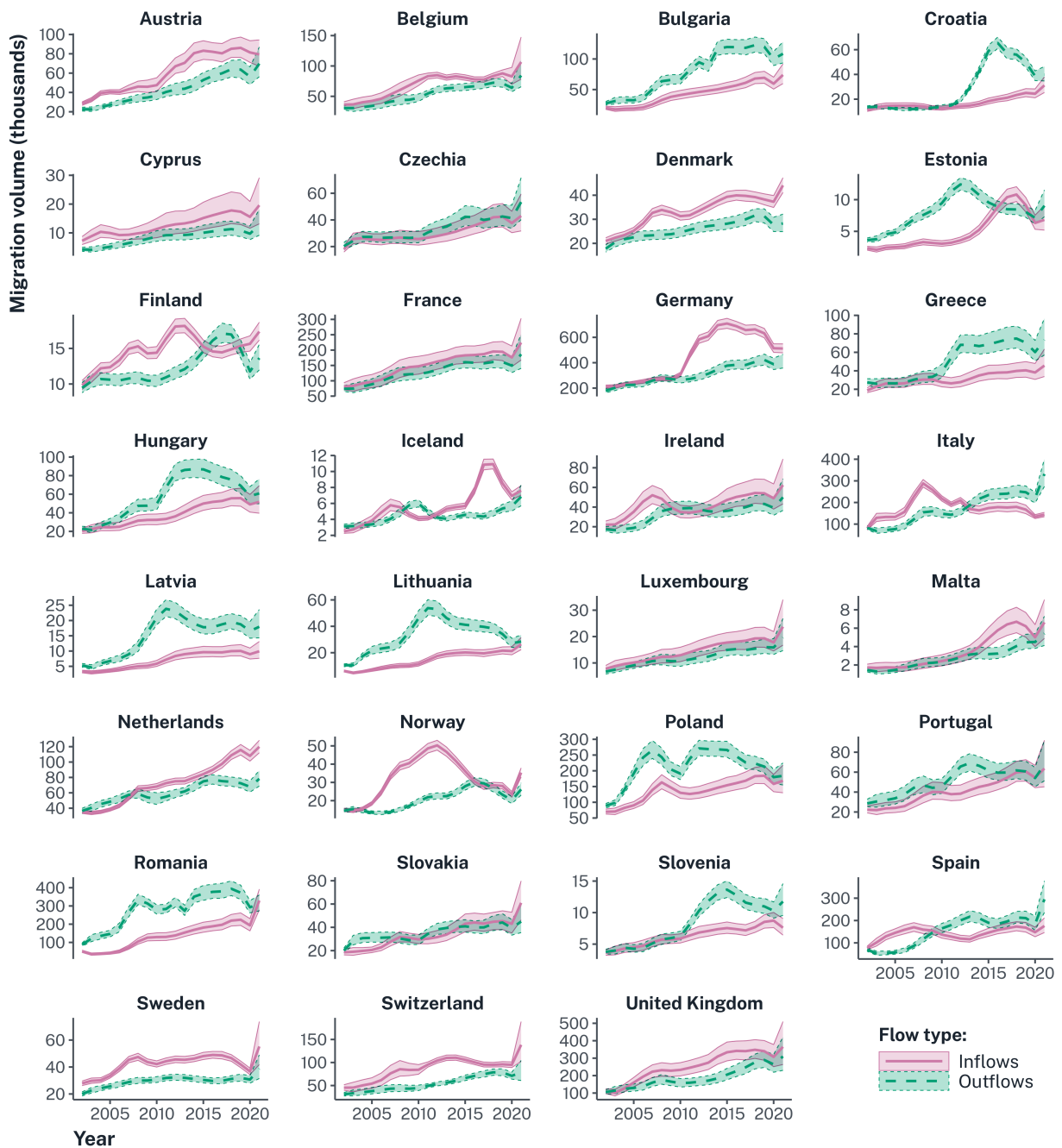
4.4.1 Data

To study international migration within Europe, I use annual, country-to-country flow estimates from 2002 to 2021 from the Human Migration Database (HMigD; Daňko, 2023; Daňko et al., 2024). Annual bilateral flows are particularly valuable for capturing temporal variation and the policy and economic shocks that occurred in the period under consideration (European Commission, 2015; Vatta, 2017). A significant advantage of the HMigD is that it corrects for misreporting and undercounting by integrating multiple data sources: Eurostat (2023a, 2023b), the United Nations Department of Economic and Social Affairs (UNDESA, 2015), the German Federal Statistical Office (Destatis, 2023), the UK's Office for National Statistics (ONS, 2020), and the IMEM and MIMOSA projects (IMEM, 2012; Kupiszewska & Wiśniowski, 2009; Raymer et al., 2013; van der Erf, 2009). These sources are harmonised using Bayesian models, incorporating bilateral flows, metadata on registration systems, and expert assessments of data reliability, and adjusting for definitional inconsistencies (eg who qualifies as an international migrant) and heterogeneous reporting practices (Daňko et al., 2024).

HMigD estimates of migration inflows and outflows for each country and year in the sample are shown in Figure 4.2. Each panel reports a country, with migration inflows shown as pink solid lines and outflows as green dashed lines; shaded ribbons denote interquartile uncertainty ranges derived from the Bayesian estimation.

Estimated Migration Inflows & Outflows in Europe by Country, 2002–2021

Bayesian estimates showing median values (lines) and interquartile ranges (ribbons).
Y-axis scales vary by country.



Source: Human Migration Database (2023).

Figure 4.2: HMigD estimates of aggregate migration inflows (in pink) and outflows (in green) in 31 European countries and over time. Posterior medians (lines) and interquartile ranges (ribbons) from the Bayesian estimation process are reported, revealing substantial cross-country heterogeneity in the volume and evolution of intra-European migration flows.

This figure illustrates the magnitude of intra-European mobility and its general upward trend, particularly in Western European countries, where inflows consistently exceed outflows. In contrast, many Eastern European countries experience persistently higher outflows, indicative of sustained emigration. The varying y-axis scales similarly reflect substantial heterogeneity in migration volumes across the 31 countries. Further, countries such as Italy and Spain display fluctuations consistent with economic cycles, policy shifts, and differences in registration systems.

4.4.2 Theory Operationalisation

In line with prior literature, gravity and MST are evaluated by translating them into sets of bilateral predictors explaining yearly flows between all country pairs from 2002 to 2021. The dependent variable is the estimated number of migrants moving from country i to country j in year t . Because the data form a complete bilateral panel and the focus is on macro-level drivers, migration flows are modelled using Poisson Pseudo-Maximum Likelihood (PPML) regressions with origin, destination, and year fixed effects. The PPML estimator is widely used in migration research due to its robustness to heteroskedasticity, ability to handle zero-valued flows, and consistency under correct mean specification (Santos Silva & Tenreyro, 2006, 2022; Weidner & Zylkin, 2021). Standard errors are three-way clustered by origin, destination, and year to account for correlated shocks and unobserved heterogeneity (Pfaffermayr, 2023).

As a point of reference, a baseline model is estimated by including only fixed effects. Its explanatory power provides a benchmark against which the added contribution of gravity and MST predictors can be assessed. The gravity and MST specifications follow the empirical applications of Ramos and Suriñach (2017) and DeWaard et al. (2012). The gravity model includes the bilateral geographic distance between country capitals and the lagged ratio of origin and destination population sizes, which remain identifiable despite fixed effects; country-specific population sizes are not considered as they are fully collinear with origin and destination fixed effects. The MST specification extends

the gravity model by including contiguity, shared language, the lagged ratio of GDP per capita between origin and destination, migrant stocks in the year 2000 (proxying network and cumulative causation effects), joint membership in the former Eastern Bloc (proxying shared geopolitical history), and lagged joint membership in the EU (proxying institutional integration). With the exception of the lagged population and GDP ratios and EU membership, all predictors are time invariant. Predictor variables and their sources are described in Appendix B.1.

Although several simulation strategies could in principle be used for Steps 2–4 of the procedure, this application uses the macro-level simulations described in Section 4.3.2. This choice reflects both theoretical and computational considerations. Gravity and MST are macro-level theories centred on bilateral linkages rather than micro-level behaviour, making a system-level simulation more appropriate. Moreover, the scale of European migration, which exceeded three million movers in 2021 alone, would render micro-level simulations computationally prohibitive. Hence, macro-level simulations represent an efficient and theoretically coherent strategy of generating theory-derived counterfactual migration systems.

4.4.3 Step 1: Regression Results

The PPML regression estimates for the baseline, gravity, and MST models are reported in Table 4.1, showing clear differences across model specifications.⁶ In the gravity model, corridors with a 10% greater geographic distance experience migration flows that are approximately 8% lower, while the coefficient of lagged population ratio is not significant. In the MST model, the distance coefficient is no longer significant; neither are factors such as lagged population ratio, contiguity, lagged GDP ratio, and lagged EU membership. Instead, the MST model underscores socio-historical linkages: migration flows between countries that share a common official language are estimated

⁶Because PPML coefficients are semi-elasticities, log-transformed continuous predictors can be interpreted in percentage terms, while effects of binary predictors correspond to $(e^{\beta} - 1) \times 100$.

to be about 67% larger than flows between otherwise comparable country pairs that do not share a common official language, while migration flows between countries with a shared geopolitical past are estimated to be about 169% larger (ie nearly 2.7 times as large) than flows between countries without such a shared history. Processes of cumulative causation also appear relevant in the MST model: holding everything else constant, a 10% increase in bilateral migrant stocks in the year 2000 is associated with a 3.9% increase in subsequent migration flows between two countries.

To assess the PPML results and detect potential model misspecification, the Ramsey Regression Equation Specification Error Test (RESET) was performed following Santos Silva and Tenreyro (2006). All models except the baseline pass the RESET with p values > 0.10 , suggesting correct model specification. Because PPML models are non-linear, a standard interpretation of the R^2 as explained variance is not possible: hence, squared correlations between fitted and observed values and an adjusted pseudo- R^2 were computed following Allison (2014) and Santos Silva and Tenreyro (2006). For the gravity and MST models, the adjusted pseudo- R^2 was computed also within origin–destination–year clusters (within adjusted pseudo- R^2). Values of these goodness-of-fit statistics show large improvements from the baseline model as more predictors are introduced. With only two predictors, the gravity model attains a squared correlation of 0.59 and an adjusted pseudo- R^2 of 0.77, consistent with previous findings on high predictive performance in gravity models of migration (Termote, 2002). The MST performs even better, attaining a squared correlation of 0.65 and an adjusted pseudo- R^2 of 0.82.

Because the dependent variable stems from a Bayesian estimation framework, the PPML models in Table 4.1 use the posterior medians of bilateral flows. To assess the sensitivity of the results to estimate uncertainty, all models were re-estimated using alternative dependent variables drawn from different posterior quantiles, exhibiting no substantive changes in the magnitude and direction of effects (see Appendix B.3, Table B.2, Table B.3 and Table B.4).

Table 4.1: Step 1: PPML regressions of different migration theories.

Model:	Baseline	Gravity	MST
<i>Predictor variables:</i>			
Distance (log)	–0.7955*** (0.1145)	–0.1959 (0.1266)	
Population ratio (lag, log)	–0.0454 (0.2756)	0.0930 (0.2845)	
Contiguity		–0.1630 (0.1538)	
Common official language		0.5117* (0.1996)	
GDP ratio (lag, log)		0.4369 (0.6576)	
Migrant stock (log)		0.3893*** (0.0578)	
Shared geopolitical history		0.9911* (0.3918)	
Shared EU membership (lag)		0.0034 (0.1251)	
<i>Fixed effects:</i>			
Origin country	✓	✓	✓
Destination country	✓	✓	✓
Year	✓	✓	✓
<i>Fit statistics:</i>			
Observations	18,600	18,600	18,600
Squared correlation	0.52252	0.58673	0.64782
Adjusted pseudo-R ²	0.72398	0.76556	0.82276
Within adjusted pseudo-R ²		0.15064	0.35789
Ramsey RESET	0.0334	0.1860	0.2032
<i>Notes: Standard errors (in parentheses) clustered by origin, destination, and year.</i> <i>Adjusted pseudo-R² is McFadden's R². RESET p values computed following Santos Silva and Tenreyro (2006). Signif. codes: †: 0.1, *: 0.05, **: 0.01, ***: 0.001.</i>			

In addition, to address potential concerns that some bilateral predictors, such as geographic distance, may have been included in the HMigD estimation process and potentially inflate the explanatory power of the PPML models, all regressions were replicated on a subsample of corridor-year observations using Eurostat flow data as the dependent variable (Eurostat, 2023b). The results of this replication appear consistent with the explanatory power of the HMigD models, indicating that high performance is not driven by statistical dependence between the outcome variable and the predictors (see Appendix B.4, Figure B.1 and Table B.5). Importantly, any undetected correlations between predictors and estimated flows would bias the subsequent simulations toward resembling the observed system, making the spatial evaluation in Steps 2–4 conservative rather than overstated. Finally, as the period under study includes the COVID-19 pandemic, which has been found to have severely disrupted both internal and international migration flows worldwide (González-Leonardo et al., 2022; Klein et al., 2024; Lerpold et al., 2023), the PPML regressions were replicated excluding the years 2019–2021: this replication shows no substantial changes in explanatory power for either theory and suggests that this period did not significantly impair the models’ performance (see Appendix B.5, Table B.6).

4.4.4 Steps 2–4: Simulation Results

The fitted values from the PPML regressions in Table 4.1 serve as inputs for Steps 2–4. Following the simulation strategy described in Section 4.3.2, 10,000 synthetic migration systems are generated for each theory and year. In each simulated system, flows are determined exclusively by the predictors of the corresponding PPML specification. The resulting set of simulations provides a distribution of gravity- and MST-specific migration systems against which the observed system can be evaluated. Steps 3 and 4 implement a statistical assessment of spatial accuracy at increasingly fine levels of resolution, starting at the system level using migration indices, then at the country level using Mahalanobis distance, and finally at the corridor level in four edge cases.

System-Level Evaluation: Migration Indices

The first level of assessment examines the similarity of observed and simulated patterns at the system level, using four migration indices: the ACV, the Gini Index, the Migration Inequality Index, and the Reciprocity Index. The results are shown in Figure 4.3, where, for each year, the distributions of simulated values are shown in blue for gravity and in orange for MST. Observed values are shown as black solid lines with interquartile uncertainty ribbons. Mean values of the simulations are marked with dotted (gravity) and dashed (MST) lines to facilitate comparison. Overlaps between simulated and observed values indicate that the model could reproduce a spatial feature under a regression uncertainty of 95%.

The results reveal systematic patterns. For migration concentration, both gravity and MST reproduce the observed levels reasonably well, albeit with some differences across indices. Gravity simulations align closely with the observed Gini Index and Migration Inequality Index, indicating that a parsimonious specification based on geographic distance and relative population size is sufficient to capture much of the overall concentration of intra-European migration. However, gravity underpredicts the ACV in some years, suggesting that gravity struggles to capture the influence of large, dominant corridors, to which the ACV is more sensitive. The MST model performs marginally better across all concentration measures and is the only specification that consistently reproduces the observed values across the three indices.

Moving to reciprocity, we observe a different pattern. The observed reciprocity values are very high, surpassing 0.40 in several years (2002, 2003, 2020), meaning that, in those years, over 40% of migrants were moving in corridors where flows occur bilaterally. Yet, the two theories predict a much more asymmetric structure, failing to reproduce the observed reciprocity in all years except 2012 and 2013, when the range of gravity-derived values includes the observed reciprocity. Overall, gravity appears to perform better than MST in reciprocity, as its simulation range covers slightly higher values.

Observed and Simulated Migration Indices, 2002–2021

Observed values in black, with median values (lines) and interquartile ranges (ribbons).
 Simulated values in blue (gravity) and orange (MST), with simulation means (lines) and entire range of distributions (ribbons). Y-axis scales differ by index.

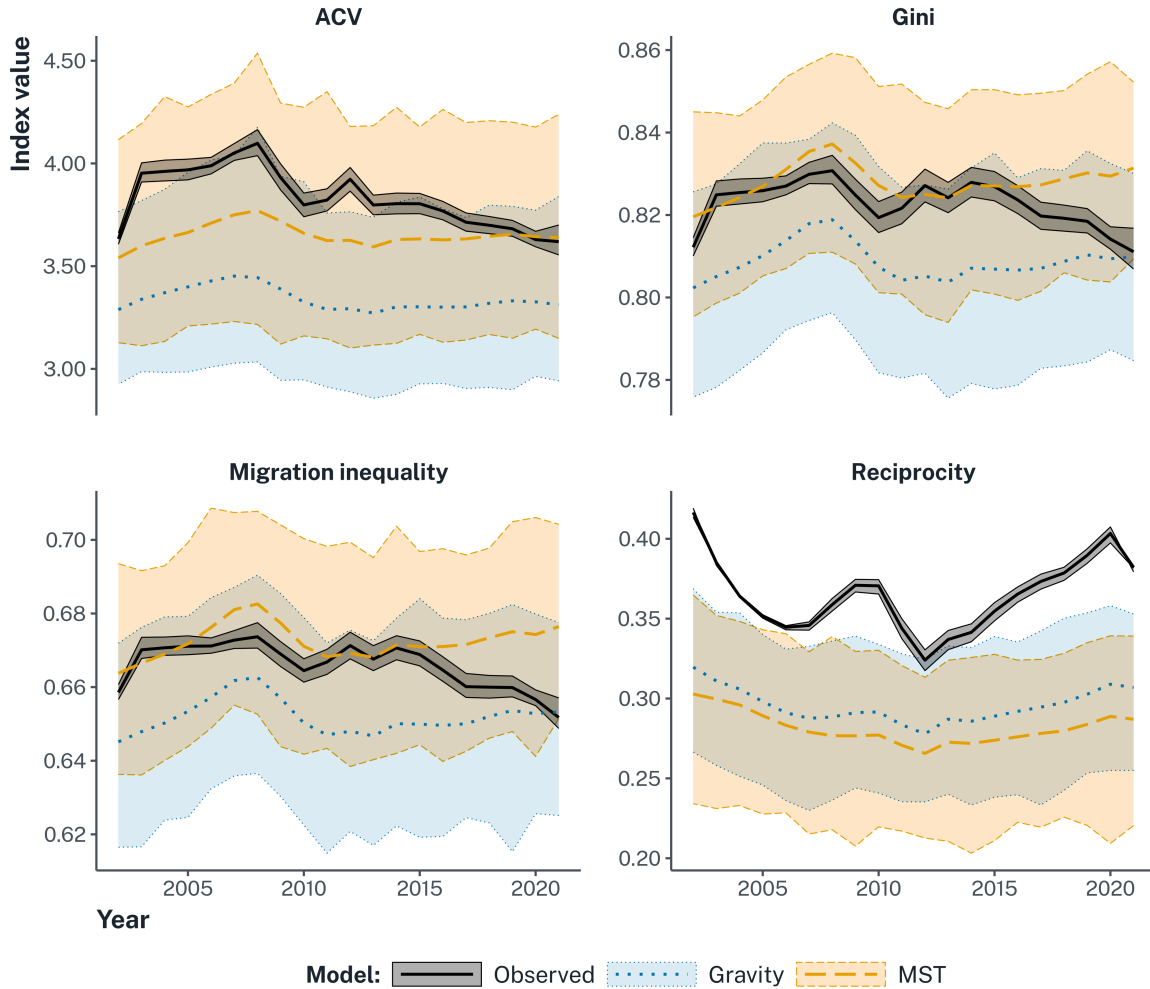


Figure 4.3: Accuracy of two explanatory models in accounting for the aggregate spatial features of the intra-European migration system over time. Spatial features measured using Aggregate Coefficient of Variation (ACV), Gini, migration inequality, and reciprocity. Observed values are reported in black, with posterior medians as solid lines and interquartile ranges as shaded ribbons. The full outcome ranges generated by 10,000 simulations per year and model, propagating the 95% regression uncertainty in the migration flows estimated in Table 4.1, are reported as shaded ribbons, in blue for gravity and orange for MST, with simulation means as dotted and dashed lines, respectively. Overlaps between the observed value and the simulation distributions indicate that the model can reproduce the observed pattern given regression uncertainty.

Country-Level Evaluation: Mahalanobis Distance in Inflows

Moving to the intermediate level of assessment, Figure 4.4 reports Mahalanobis distances between observed and simulated inflows for each country and year, separately for gravity (blue) and MST (orange). This measure captures how unusual each country's observed inflow profile is relative to what each model predicts.

Several insights emerge. First, MST achieves a broadly better performance (ie lower distances) for most countries, in particular for Austria, Germany, Hungary, Ireland, Latvia, Poland, Portugal, and Switzerland. Second, gravity performs relatively better in Nordic countries (Denmark, Finland, Norway, and Sweden), France and the Netherlands pre-2008, and Luxembourg pre-2010. Third, the magnitude of Mahalanobis distances reveals where and when flows are hardest to predict, regardless of theory. In countries often characterised as peripheral within the European migration system (eg Croatia, Cyprus, Czechia, Estonia, Finland, Greece, Lithuania, Malta, Romania, Slovakia, and Slovenia), observed flows tend to lie far from both sets of simulations, as evidenced by the larger Mahalanobis distance values ($D^2 > 4.00$). Fourth, some countries exhibit sharp spikes over specific periods, such as Estonia in 2016–2017, Iceland in 2017, Lithuania, Romania, and Slovakia during the years affected by the COVID-19 pandemic (2020–2021), and Italy between 2007 and 2014, suggesting the influence of unmodelled shocks beyond what could be controlled for by the country and time fixed effects.

Corridor-Level Evaluation

At the most granular level of assessment, Figure 4.5 examines the spatial accuracy of the two theories for four example corridors: Austria to Germany, Cyprus to Greece, Estonia to Finland, and Finland to Estonia. The four corridors were chosen as they represent edge cases that best illustrate when and where different theoretical frameworks may be appropriate.

Distance Between Observed and Simulated Inflows by Country, 2002–2021

Mahalanobis distance between observed and simulated inflows, with gravity values in blue and MST values in orange. Y-axis scales vary by country.

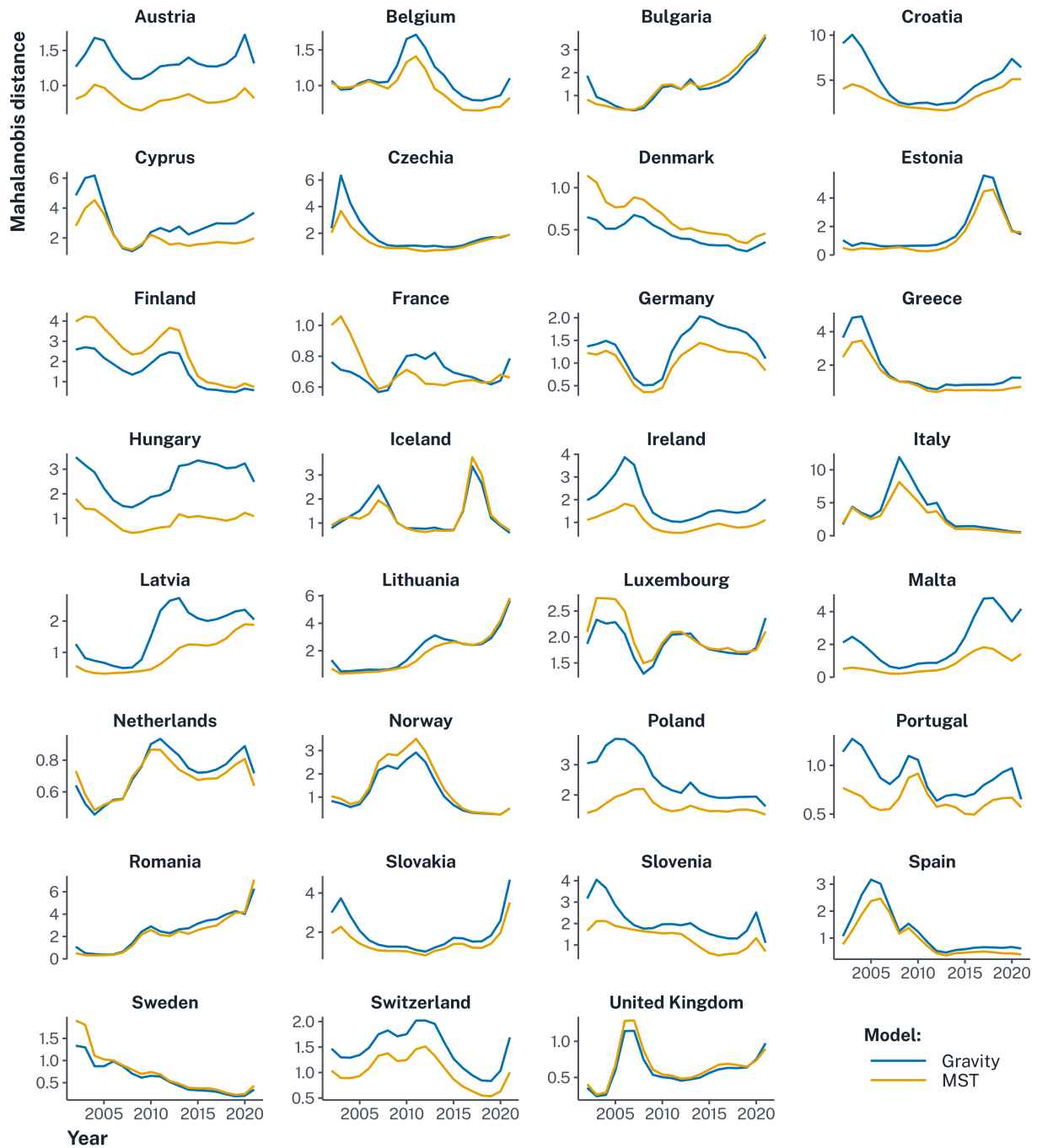


Figure 4.4: Distance between observed and simulated inflows in the intra-European migration system by country and over time. Mahalanobis distances were computed by extracting vectors of inflows for each country and year from the observed migration system, and comparing these vectors with 10,000 equivalent vectors stemming from the model-based simulations. This process was conducted separately for the two models, gravity (blue) and MST (orange). Lower Mahalanobis distance values indicate a closer match between simulations and observed migration patterns.

Observed flows are reported as black solid lines with interquartile uncertainty ribbons; simulated flow ranges as blue (gravity) and orange (MST) ribbons, with simulation means as dotted (gravity) and dashed lines (MST). Overlaps between observed values and simulated distributions indicate that the observed flow volume is consistent with what a theory would generate, given a regression uncertainty of 95%.

In the corridor from Austria to Germany, which involves two contiguous EU countries that share a common official language, both gravity and MST perform well, as the observed flows always fall within the range of simulated values. Conversely, in the case of Cyprus to Greece, which involves two countries relatively distant geographically but connected by historical ties and a common language, only the MST model performs well, as the gravity simulations seldom touch the observed range. Finally, the two reciprocal corridors, Estonia to Finland and Finland to Estonia, could not be easily replicated by either theory, even though MST achieves an overall better performance. Both corridors exhibit sudden spikes in migration volume in the periods 2010–2015 (Estonia to Finland) and 2015–2020 (Finland to Estonia), which could not be accounted for by the gravity and MST predictors.

As a complementary diagnostic, Appendix B.6 reports scatterplots of observed *vs* simulated flows for all corridor–year observations (see Figure B.2 for gravity, Figure B.3 for MST, and Figure B.4 for the four exemplary corridors). Three main patterns emerge. First, there is no evidence of systematic temporal drift in model performance: deviations from the diagonal (denoting perfect agreement between simulated and observed values) remain broadly stable over time. Second, while MST performs slightly better than gravity for several regional corridors, the relative performance of the two models is otherwise heterogeneous and does not follow a consistent regional hierarchy. Lastly, one pattern appears robust across models: migration corridors connecting two Eastern European countries, two Northern European countries, and from Northern to Eastern European countries are systematically poorly predicted, with the two theories achieving the lowest spatial accuracy.

Observed and Simulated Flows in Selected Corridors, 2002–2021

Observed values in black, with median values (lines) and interquartile ranges (ribbons). Simulated values in blue (gravity) and orange (MST), with simulation means (lines) and entire range of distributions (ribbons). Y-axis scales differ by country.

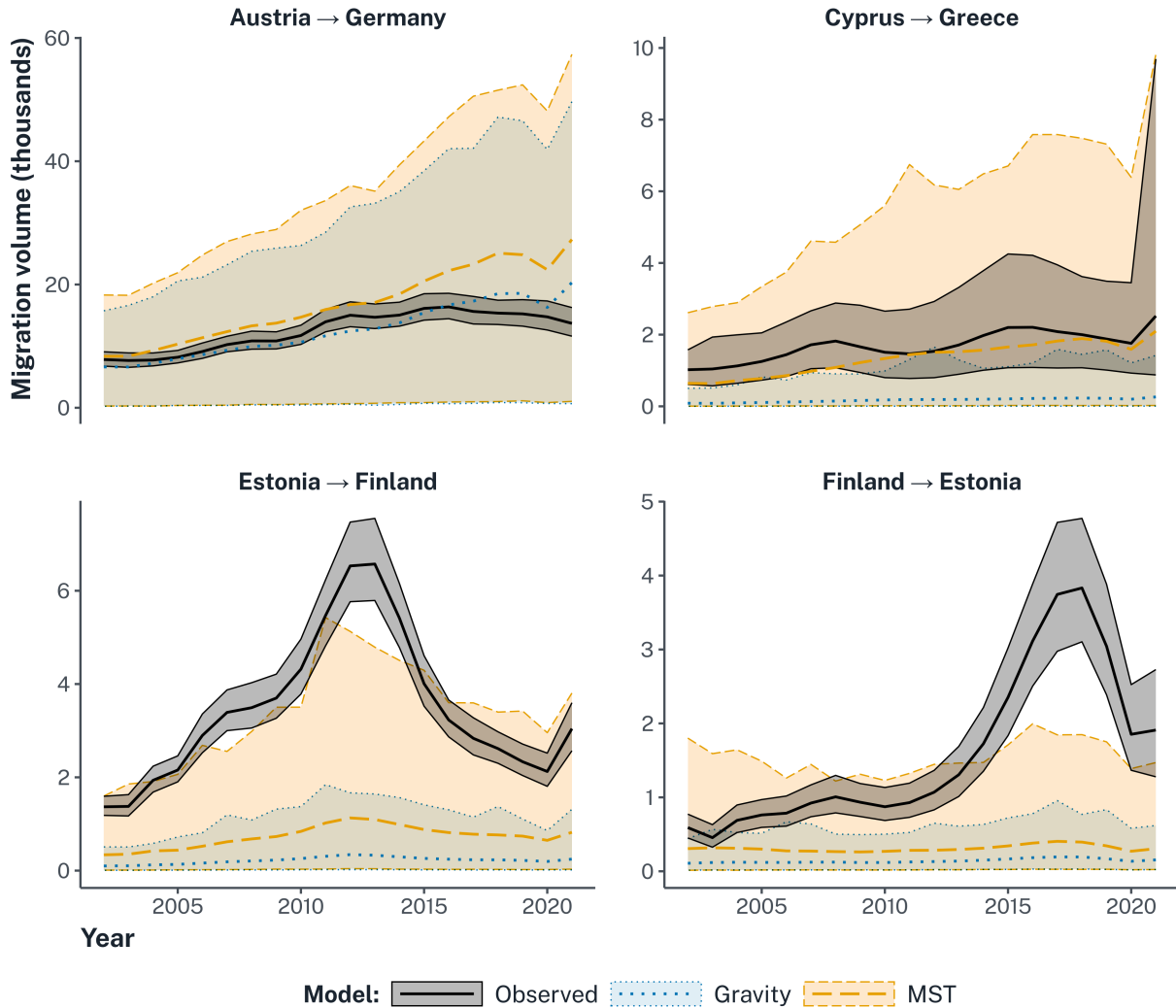


Figure 4.5: Accuracy of two explanatory models in accounting for the migration flows occurring in four exemplary corridors over time. For each of the four corridors, observed values are reported in black, with posterior medians (solid lines) and interquartile ranges (shaded ribbons). The full outcome ranges generated by 10,000 simulations per corridor, year, and model, propagating the 95% regression uncertainty in the migration flows estimated in Table 4.1, are reported as shaded ribbons, in blue for gravity and orange for MST, with simulation means as dotted and dashed lines, respectively. Overlaps between the observed value and the simulation distributions indicate that the model can reproduce the observed pattern given regression uncertainty. The four corridors were selected as illustrative contrasts: (1) flows in the first corridor, Austria to Germany, are well reproduced by both models; (2) flows in the second corridor, Cyprus to Greece, are well reproduced by MST only; (3-4) flows in the last two corridors, Estonia to Finland and Finland to Estonia, are not well reproduced by either model.

4.5 Discussion

This chapter set out to evaluate migration theories not only by their explanatory power, but also by their ability to reproduce the spatial organisation of migration flows. The simulation-based spatial test applied to intra-European migration between 2002 and 2021 produces a clear result: theories that perform well by conventional standards may systematically misrepresent how migration is organised across space. Both gravity and MST achieve strong explanatory power, yet fail to reproduce the high reciprocity of European flows and the distinctive migration profiles of Eastern and Northern European countries. These shortcomings are not uniform—they vary across spatial scales, time periods, and regional contexts—but are consistent enough to indicate that standard goodness-of-fit measures do not guarantee spatial accuracy.

The core contribution of this chapter thus lies in demonstrating that explanatory power and spatial accuracy constitute analytically distinct dimensions of theory-testing. In this application, MST explains a larger share of variance in bilateral flows than gravity, yet both theories generate simulated migration systems that diverge from reality in fundamental ways. Reliance on regression-based goodness-of-fit measures alone would have supported a straightforward conclusion that MST provides the superior explanation of intra-European migration flows.

Yet, the spatial evaluation shows that such a conclusion would be incomplete: despite strong explanatory power, even MST may systematically misrepresent how migration is organised across space and time. If we understand migration patterns as the empirical imprints of underlying mechanisms, these misfits are not minor deviations, but signal that the theories differ in important ways from the pattern-generating processes of the observed system.

Building on the regression results, the simulation-based analysis evaluates spatial accuracy at increasingly fine levels of spatial resolution. At the system level, the two theories exhibit nuanced but distinct patterns of spatial fit. For migration concentration,

both theories perform well overall, with the MST model performing marginally better. By contrast, for reciprocity, both theories fail to match the high levels observed in most years: gravity performs relatively better than MST, whose simulations imply more asymmetric exchange patterns.

Importantly, the reciprocity misfit should be interpreted in light of Europe's distinctive institutional setting. Freedom of movement within the EU, complemented by bilateral mobility agreements with non-EU countries such as Switzerland, creates unusually symmetric conditions for cross-border mobility across a large set of countries, which facilitate not only permanent relocation but also short-term, temporary, and circular forms of migration. Such forms of mobility are common in labour migration and international student mobility and can generate high levels of reciprocal exchange even when long-term settlement remains uneven. These dynamics are prevalent in Europe, but only weakly theorised by gravity and MST, helping to explain why both models systematically underestimate reciprocity at the system level.

Moving to the country level, the spatial tests reveal further heterogeneity in model performance across countries and time. Several countries exhibit particularly volatile time trends, which indicate the interference of sudden shocks. An example of this is the COVID-19 pandemic, which affected the gravity and MST performance in the case of Lithuania, Romania, and Slovakia. Given the extraordinary policy interventions that were implemented during this period, curtailing both internal and international migration, the COVID-19 pandemic can be understood as a stress test for the models, where misfits likely reflect unpredictable policy shocks, rather than theoretical shortcomings.

Looking at other country cases, MST generally outperforms gravity across much of Central Europe, where historical, geopolitical, and institutional linkages appear to play a stronger role. At the same time, MST performs poorly in Northern Europe. This discrepancy is instructive. MST emphasises hierarchical core–periphery relations rooted in unequal political and economic ties: while such mechanisms are highly

relevant at a global scale, within Europe, and especially Northern Europe, migration is organised more through multidirectional exchanges. In this sense, Europe as a whole can be seen as a global migration ‘core’. Nordic countries exhibit further distinctive institutional and social features, such as extensive welfare systems, regulated labour markets, and high levels of institutional integration, that set them apart from other regional clusters and are not explicitly incorporated into either gravity or MST.

Theoretical shortcomings become visible also at the level of migration corridors. A striking example is provided by the reciprocal migration connecting Estonia and Finland. Both gravity and MST fail to reproduce the sharp increase in migration from Estonia to Finland between 2010 and 2015, followed by a subsequent rise in migration in the opposite direction between 2015 and 2020. Existing research links the Estonia–Finland pattern to post-enlargement labour mobility, student migration, and later return or circular migration facilitated by geographic proximity and institutional integration (Anniste & Tammaru, 2014; Jakobson et al., 2012; Toomistu et al., 2024). While such dynamics of return and circular migration naturally generate high reciprocity at the system and corridor level, they are not explicitly captured by either theory. More generally, the diagnostic scatterplots in Appendix B.6 show that corridors within certain regions, in particular within Eastern Europe, within Northern Europe, and from Northern to Eastern Europe, are systematically poorly reproduced. This shared misfit points to common theoretical blind spots in accounting for the specific regional clusters of intra-European migration.

The present empirical application cannot directly test the role of return and circular migration in shaping intra-European flows, as it relies on aggregate bilateral data that do not distinguish between first-time moves, return migration, or repeated cross-border mobility. As a result, specific life-course mechanisms (such as temporary labour migration followed by return) cannot be isolated, and competing explanations of reciprocity at the corridor level cannot be tested. Importantly, this limitation pertains to the data and simulation design of this specific application, rather than to the proposed

method itself. The same procedure could be implemented using individual-level data and micro-level simulations to evaluate theories operating at lower levels of analysis, such as network theory or the new economics of labour migration.

A further limitation concerns geographical scope. By focusing on intra-European international migration, the analysis excludes some of the world's most prominent corridors, including flows from the Global South to the Global North. At the global level, migration systems are characterised by stronger asymmetries in income, legal status, and mobility rights—conditions under which the hierarchical expectations of MST may find greater empirical purchase. Applications beyond Europe could therefore assess whether MST performs better in explaining historically asymmetric and hierarchical systems. Similarly, future work could exploit the method's portability and consider internal migration cases, evaluating migration theories relevant to within-country settings.

The broader implication of this chapter is that spatial accuracy deserves consideration alongside explanatory power when evaluating migration theories. The misfits documented here—particularly the reciprocity shortfall and the regional clustering of errors—are not minor residuals but diagnostic signals that point toward processes that existing theories do not capture. Making such signals visible, and quantifying their magnitude, is what the simulation-based framework is designed to do.

5| The Interdependence of Migration: How Relational Structures Organise Movement*

Abstract

Migration theory emphasises that mobility decisions are embedded in relational structures, such as interpersonal networks and organisational pipelines, and are therefore interdependent. Yet, most quantitative models treat migration outcomes as determined exclusively by the observable attributes of people and places and conditionally independent once such attributes are controlled for. I address this tension by capturing the effect of relational structures through the interdependence traces they leave in aggregate data, using a recently proposed statistical network model designed specifically for human mobility. Comparing degree student mobility with Erasmus exchanges across 29 European countries—two cases involving similar populations but operating under distinct institutional architectures—I show that relational structures shape destination choices in ways that differ systematically across contexts and that standard models cannot capture. Accounting for interdependence also alters conclusions about conventional predictors, including geographic distance. Modelling relational structure is thus not merely a technical refinement but a necessary step toward aligning empirical migration research with its theoretical foundations.

*Material from this chapter has been used to prepare an article co-authored with Per Block, currently under review.

5.1 Introduction

Sociological research on migration has long recognised that individual decisions to move, both internally and internationally, are shaped not only by personal characteristics and contextual conditions, but also by meso-level drivers such as interpersonal networks, institutional partnerships, and organisational channels (Bakewell, Engbersen et al., 2016; de Haas, 2010; Faist, 1997, 2000; Gold, 2005). These meso-level drivers link actors to actors and actors to places, thus building *relational structures*: migrants follow paths established by predecessors, rely on social networks for information and support, and move through channels created by institutions, firms, universities, and recruitment agencies (Cranston et al., 2018; Findlay & Li, 1998; Garip & Asad, 2016; Haug, 2008; Palloni et al., 2001; Xiang & Lindquist, 2014).

A central implication of the relational perspective is that migration decisions are *interdependent*: one person's decision to move is affected by, and in turn affects, the mobility of others. As implied by social mechanisms such as chain migration and cumulative causation, interdependence arises as realised migration events accumulate over time and earlier movers shape the opportunity structures faced by others, so that migration along a certain corridor affects the future attractiveness of that same corridor (Bakewell & Jolivet, 2016; de Haas, 2010; Engbersen, Snel & Horst, 2016; Massey, 1990). Interdependence also arises when organisations and institutions establish migration routes and encourage, or even force, mobility along them, so that the moves of different individuals are co-determined (Findlay & Li, 1998; Wittek, 2021; Xiang & Lindquist, 2014). In both instances, migration flows are coordinated, channelled, or constrained in ways that cannot be reduced to individual or contextual attributes alone.

Despite its theoretical centrality, interdependence remains empirically elusive in migration research. The relational structures that generate it are rarely observed comprehensively or comparably at scale and in different contexts. Therefore, most quantitative models of migration, particularly those based on aggregate flows, assume

that migration decisions and flows are conditionally independent once observable factors are controlled for. This assumption of independent observations is seldom theoretically justified; rather, it reflects a modelling convenience that enables the implementation of standard regression techniques (Aksoy & Yıldırım, 2024; Huang & Butts, 2023; Leal, 2021; McMillan, 2024).

The cost of this modelling strategy is an open question. If interdependence exerts only a marginal influence on aggregate flows, or operates orthogonally to conventional predictors, treating migration decisions as independent would be largely inconsequential for inference. If, however, interdependence strongly organises migration flows, independence-based models will have limited explanatory power; where it further correlates with macro-level conditions, they will yield biased estimates, misattributing explanatory weight to conventional predictors. This chapter puts this question to an empirical test, asking: (1) *To what extent does interdependence structure migration flows beyond conventional predictors?* (2) *Does ignoring interdependence distort estimates of those predictors?*

To this end, I rely on a statistical network model, the Endogenous Log-linear Model (ELMo), designed to estimate interdependence directly from aggregate data, even when the relational structures generating it are unobserved (Block et al., 2022; Hoffman & Block, 2026). The ELMo extends a multinomial logit framework with endogenous parameters that capture different forms of interdependence. I focus on origin interdependence, which links the destination choices of individuals from the same origin, and dyadic interdependence, which links migration flows occurring in opposite directions in a migration corridor. Estimating these endogenous parameters alongside conventional predictors allows me to quantify interdependence directly and examine whether ignoring it distorts inference about established migration drivers.

To better understand whether the proposed modelling strategy captures genuine features of the migration process, I apply it to a setting that provides strong inferential leverage: International Student Mobility (ISM) in Europe. This case enables a controlled

comparison between degree mobility, where students pursue full degrees abroad, and Erasmus credit mobility, where exchanges are typically shorter and occur through formal inter-institutional partnerships. This means that the two mobility types involve similar populations moving between the same countries in the same year (2019), but differ sharply in their organising principles. Degree mobility is largely shaped by individual choice and embedded in social processes; Erasmus mobility is channelled through inter-institutional agreements. If the ELMo captures genuine interdependence rather than unmodelled confounding, the origin and dyadic parameters should diverge systematically across these institutionally distinct systems.

I find that interdependence is empirically meaningful and varies across ISM types. I observe significant origin interdependence in both degree and Erasmus mobility, with students from the same origin concentrating on the same destinations; however, effects are substantially stronger in degree mobility, where origin interdependence accounts for 40% of relative explanatory importance, exceeding any conventional predictor. Dyadic interdependence diverges across ISM types: while weakly positive in degree mobility, it is significantly negative in Erasmus mobility, indicating systematic asymmetry in flows despite the programme's stated principle of reciprocal and balanced exchange. This points to previously overlooked institutional motivations to participate in the Erasmus programme. Accounting for interdependence also reshapes estimates of conventional parameters: effects that appear substantive in standard multinomial logit models, such as geographic distance, diminish or reverse once interdependence is introduced.

These findings contribute to migration research in three ways. Theoretically, this chapter clarifies how the relational nature of migration can manifest in distinct forms of interdependence (origin and dyadic) that structure collective outcomes in ways that cannot be reduced to individual or contextual attributes. Methodologically, the analysis demonstrates that interdependence can be estimated from aggregate flow data, even when the underlying relational structures are unobserved, by operationalising it

as endogenous parameters in a statistical network model. Substantively, the chapter shows that modelling interdependence explicitly can reshape conclusions about the forces organising human movement. Taken together, the results suggest that modelling interdependence is not merely a technical refinement, but a necessary step toward aligning empirical analysis with core sociological insights about the embeddedness of migration processes.

The remainder of the chapter develops this argument. In Section 5.2, I introduce the theoretical framework, distinguishing origin and dyadic interdependence and situating them relative to the literature on migration drivers. In Section 5.3, I outline the research questions and analytical strategy. In Sections 5.4, 5.5, and 5.6, I describe the ISM case, data, and methods. In Section 5.7, I present the empirical results and in Section 5.8 their discussion.

5.2 Migration Drivers and Relational Interdependence

Migration research has identified a wide range of factors, commonly known as *drivers*, that enable, restrict, and motivate internal and international migration. Migration drivers represent the conditions influencing the propensity to move and describe ‘what exists’ in a given context: resources, opportunities, and constraints that make moving possible, attractive, or costly. Conventionally, researchers distinguish between drivers operating at the micro, meso, and macro level, with the understanding that drivers at different levels typically interact with one another (Czaika & Reinprecht, 2022).

At the micro level, migration drivers include individual characteristics, aspirations, capabilities, and life-course transitions (de Haas, 2021). At the meso level, drivers encompass social and organisational structures such as interpersonal networks, firms and institutions, and recruitment intermediaries (Cranston et al., 2018; Czaika & Reinprecht, 2022; Faist, 1997, 2000; Findlay & Li, 1998; Xiang & Lindquist, 2014). At the macro level, drivers refer to the characteristics of origins and destinations, such

as geographic distance between places, economic and labour market conditions and differentials, demographic pressures, and politico-institutional linkages and legacies (Massey et al., 1993; Mayda, 2010).

This chapter focuses on meso-level drivers, as they are fundamentally *relational*. Unlike micro-level drivers, which describe individual attributes, or macro-level drivers, which describe characteristics of places, meso-level drivers concern the connections between people and people, and between people and places, across origins and destinations (Faist, 1997, 2000; Findlay & Li, 1998; Gold, 2005). Interpersonal networks, institutional partnerships, organisational channels, and recruitment intermediaries are relational in this sense: they link actors to one another and to locations, mediating between individual decisions and the broader contexts in which migration occurs. Because of this, meso-level drivers constitute *relational structures*: durable configurations of ties, channels, and pathways through which migration is organised.

The defining feature of relational structures is that they produce *interdependence* among migration decisions. Through relational structures, individual choices to move become contingent on one another. Interpersonal networks link prospective migrants to diaspora communities abroad; firms and recruitment agencies channel workers to specific locations; institutional agreements steer students toward partner universities. In each case, the decision of one person to move is shaped by the decisions—past and present—of others.

Migration interdependence, discussed in some contexts as feedback (cf Bakewell, Engbersen et al., 2016; de Haas, 2010; Leal, 2021), implies that migration decisions and, by extension, aggregate flows are connected beyond what individual, origin, and destination attributes can explain. When someone migrates, the informational environment, social support infrastructure, and institutional channels available to others are altered. These alterations operate both synchronically, as contemporaneous movers may influence one another's destination choices through shared networks and common institutional constraints, and diachronically, as earlier moves reshape

the conditions under which subsequent migration occurs. These dynamics give rise to observable regularities in the spatial organisation of migration beyond micro- and macro-level conditions: persistence and path dependence in migration routes, concentration of flows toward specific destinations, and systematic imbalances in migration corridors (Bakewell, Engbersen et al., 2016; de Haas, 2010).

Depending on how relational structures link migration decisions between individuals and places, they generate distinctive forms of interdependence. Here, I focus on two such forms: origin and dyadic interdependence.

5.2.1 Origin Interdependence

Origin interdependence describes the tendency for destination choices among individuals from the same origin to be correlated, beyond what shared individual characteristics or exposure to the same macro-level conditions can explain. Where this correlation is positive, migrants from a given origin will concentrate on the same destinations. Where it is negative, they will spread more evenly than expected across the available migration routes.

Several strands of migration research make a theoretical case for positive origin interdependence. The most prominent stems from the study of peer and social influence and migrants' interpersonal networks. As individuals from a particular origin settle at a specific destination, they often seek family reunification and encourage their relatives to join them in the destination (Fussell, 2010; Kofman et al., 2022). Earlier movers typically provide prospective migrants with information about opportunities, logistical assistance, and social anchoring in the destination, thereby lowering the costs and risks of moving along that same corridor, in what is known as direct social facilitation (Garip & Asad, 2016; see also Haug, 2008; Liang et al., 2008; Massey & Espinosa, 1997; Palloni et al., 2001). As diaspora communities continue to grow in the destination, network externalities accumulate: a broader infrastructure of co-ethnic or co-national businesses, community organisations, and specialist services emerges, benefitting

newcomers regardless of whether they hold direct personal ties to earlier movers (Garip & Asad, 2016). In this context, the destination becomes progressively more attractive, especially in comparison to alternatives that might offer similar conditions, at least in terms of macro-level indicators, but lack comparable social facilitators and support networks upon arrival. Normative influence adds a further layer to this process: when migration to a particular destination becomes common within an origin community, that destination may be perceived as the 'go-to' option anytime someone considers relocating. Migration itself may become institutionally and socially expected, in what the literature terms a 'culture of migration', even when the act of moving may no longer be 'beneficial' according to a narrow cost-benefit comparison that abstracts from social relations (Garip & Asad, 2016; Heering et al., 2004; Kandel & Massey, 2002).

Positive origin interdependence may also be generated by organisational forces. When firms, recruitment agencies, or educational institutions systematically hire, transfer, or enrol people from the same origins to the same destinations, the result would be a similar concentration pattern. However, in this case concentration is driven not by interpersonal networks but by organisational channelling. A typical case where this has been theorised to occur is through multinational corporations, where international migration occurs because of assigned expatriation: in such cases, employees are transferred along corporate hierarchies that reflect organisational needs rather than individual preferences (Andresen et al., 2014; Przytuła, 2015). Indeed, assigned expatriations often follow established intra-firm pathways, for instance from specific national subsidiaries to headquarters or regional hubs (Ewers & Dicce, 2018; Wittek, 2021). Similarly, recruitment agencies may produce positive origin interdependence by repeatedly sourcing workers with particular profiles from the same origins and placing them in the same destinations, often to satisfy labour demands in specific sectors. This effectively steers migrants along pre-existing channels rather than across the full set of possible alternatives (Cranston et al., 2018; Debonneville, 2021; Findlay & Li, 1998; Muranaka, 2022; Xiang & Lindquist, 2014).

Although less established in empirical studies, origin interdependence may also be negative. Stouffer's model of intervening opportunities suggests that when migration along a particular corridor approaches saturation, prospective migrants may redirect towards less congested alternatives, resulting in an even distribution of flows across the system (Stouffer, 1960). Similar patterns have been theorised by research on migrant competition and multilateral resistance to migration (Batista & McKenzie, 2023; Beine et al., 2025; Bertoli & Fernández-Huertas Moraga, 2013). Further, the same social structures that generate positive origin interdependence have been hypothesised to operate in reverse: if settled migrants experience deteriorating labour market conditions, restrictive policy environments, or hostile reception in the destination, they may try to systematically redirect prospective movers to other locations (Engbersen, Snel & Esteves, 2016; Fussell & Massey, 2004; Paredes-Orozco, 2019). In both cases, the emerging pattern would be that of an even spread of migration flows across available corridors.

5.2.2 Dyadic Interdependence

Dyadic interdependence describes the tendency for flows occurring in both directions of a migration corridor to be systematically related to each other, beyond what the origin and destination attributes can explain. Where this tendency is positive, flows in one direction of the corridor are more likely to occur when counterflows are also present, in a form of migration reciprocity. Where it is negative, flows in one direction are less likely to occur in the presence of counterflows, in a form of structural asymmetry.

In the migration literature, dyadic interdependence has received less theoretical attention than origin interdependence, but several organisational mechanisms are relevant to its emergence. Firms, corporations, and institutions enter bilateral partnerships where they 'exchange' or lend one another workers and professionals: in such cases, exchanges in one direction are systematically, and often formally, associated with exchanges (or lack thereof) in the opposite direction (Adhvaryu et al., 2024; Wittek,

2021). Additionally, organisations operating internationally may transfer workers back-and-forth between the countries in which they operate. Similarly, universities and other Higher Education Institutions (HEIs) often establish inter-institutional agreements, such as the Erasmus programme in Europe, to promote both student and staff mobility along bilateral corridors (Findlay, 2010; Recchi, 2015).

Both instances could result in either positive—if organisations seek to establish balanced, reciprocal exchanges—or negative dyadic interdependence—if organisations seek to fill specific skill shortages or pursue asymmetric goals, such as prestige acquisition, capacity building, or broader internationalisation strategies (Findlay, 2010; Findlay et al., 2012; Teichler, 2017). Importantly, whenever migration is triggered or promoted by organisational forces, individual choice sets in destination selection may be affected. Individuals may have the aspiration and capability to migrate, but the destinations they can realistically access depend on organisational decisions that lie outside their control: for instance, based on which firm subsidiaries are strategically made available, or which partner universities are accredited and recognised as viable by their home institutions.¹

5.2.3 Relational *vs* Macro-level Drivers

Before examining how relational and macro-level drivers interact, it is important to clarify what distinguishes them. Macro-level drivers include not only characteristics of individual origins and destinations but also dyadic attributes of country pairs: geographic distance, colonial history, linguistic similarity, trade relations, and visa agreements. These bilateral characteristics describe conditions that may facilitate or impede migration, but they are not themselves relational structures. Relational

¹A further case of positive dyadic interdependence is return migration. Individuals who have migrated once are more likely to migrate again, typically to their place of origin, with which they have maintained social and institutional ties (R. King, 2025; Waldorf, 1995). Through return migration, movers traverse the same corridor in opposite directions, generating reciprocal flow patterns. Strictly speaking, this represents within-person rather than between-person interdependence; however, in cross-sectional flow data, return moves are observationally indistinguishable from reciprocal moves by different individuals.

structures, by contrast, are constituted by actual ties between actors. The distinction matters because bilateral macro-level characteristics are relatively stable and exogenous to migration flows, whereas relational structures are endogenous: they are built, reinforced, and sometimes dissolved through the migration process itself.

Although analytically distinct, relational structures are often interrelated with macro-level drivers: prospective migrants know compatriots abroad who moved there *because* of economic conditions, and organisations move workers to other locations *because* those specific origins and destinations are compatible with the economic interest of the organisation. Thus, relational structures may initially reinforce the influence of macro-level drivers. However, once established, social and organisational ties will endure beyond the macro-level conditions that prompted them, sustaining migration even when those conditions change.

For social ties, this was formulated explicitly by Garip and Asad (2016): in the first stage, migrants may relocate to follow economic opportunities. In the next stage, the initial migrants transmit information, offer support, and help lower the perceived risks of moving to the same place for individuals in their origin. Over time, as this specific destination becomes the socially expected place to migrate to, the relative propensity to move there is heightened and likely to persist even if the destination conditions deteriorate or if alternative locations become more attractive, in a form of path dependence. This logic underpins well-known mechanisms of chain migration and cumulative causation, whereby individuals from the same origin repeatedly follow similar routes to common destinations, eventually reaching a critical mass after which flows become self-sustaining (Bakewell & Jolivet, 2016; Engbersen, Snel & Horst, 2016; MacDonald & MacDonald, 1964; Massey, 1990; Massey et al., 1993; C. A. Price, 1963).

A similar argument applies to the organisational cross-border connections that often underlie high-skilled migration. Even when the initial establishment of such connections follows economic opportunities, these connections can persist after macro-level conditions change, endogenising migrant flows. For example, prior to 2020,

cross-border movement between the United Kingdom (UK) and mainland Europe was facilitated by shared membership in the European Union (EU) single market. Following the UK's exit from the EU, the legal and institutional environment governing mobility changed substantially, yet firms and HEIs implemented strategies to manage migration inflows and retain employment and enrolment post-Brexit (Papatsiba & Marginson, 2025; Sumption, 2019, 2022). Because dissolving established organisational arrangements entails administrative costs, contractual renegotiation, and coordination challenges, organisations have further incentives to maintain existing cross-border channels.

Therefore, relational structures, both social and organisational, decouple from the macro-level conditions that led to their establishment, generating emergent migration systems that cannot be reduced solely to their contextual origins.

5.2.4 Existing Approaches to Modelling Migration Interdependence

As discussed so far, there is ample theoretical support for the claim that migration decisions are interdependent and shaped by relational structures. This position, however, poses distinct but related empirical challenges.

Most quantitative migration research is organised around regression frameworks in which either individual migration decisions (at the micro level) or aggregate migration volume (at the macro level) serve as the outcome, while migration drivers serve as the predictors (Boyle et al., 1998; Garip, 2012; Massey et al., 1993). In empirical micro-models of migration, identifying interdependence requires detailed measurement of the relational structures through which it is expected to operate: complete interpersonal networks, corporate recruitment pipelines, or institutional agreements and their implementation. Where such data exist, interdependence can be documented convincingly. Studies drawing on rich individual and longitudinal data show that social capital and network variables are among the strongest predictors of migration in China (Liang et al., 2008; Zhao & Qu, 2022), Congo and Senegal (Toma &

Vause, 2014), Honduras (Thomas et al., 2026), Mexico (Dolfin & Genicot, 2010; Garip & Asad, 2016; Massey & Espinosa, 1997; Palloni et al., 2001; Paredes-Orozco, 2019), Morocco (Heering et al., 2004), and Nepal (Bohra & Massey, 2009). However, such fine-grained data are difficult to obtain at a large and comparative scale, particularly for international migration involving multiple countries and corridors. As a result, the most compelling micro-level evidence for interdependence comes from settings where intensive survey data collection is feasible, primarily from the perspective of specific bilateral corridors or single-origin communities, such as the Mexico–US stream. Given that interdependence may emerge also through the presence of competing corridors, such as the one from Mexico to Spain, questions related to its influence at a broader scale remain largely unaddressed.

In empirical macro-models of migration, analytical strategies are more constrained. In models of this kind, bilateral migration flows are treated as conditionally independent given the included predictors. The flow from origin A to destination B is assumed to be unrelated to the flow from A to destination C or from B to A once observable contextual characteristics are controlled for. This assumption is methodological rather than theoretical: by construction, it allows estimation of flows under conventional regression assumptions, in particular the independence of observations assumption (Aksoy & Yıldırım, 2024; Beine et al., 2016, 2025; Bertoli & Fernández-Huertas Moraga, 2013). However, it also rules out the interdependence that migration theory identifies as central to the spatial organisation of movement. When interdependence is acknowledged, it is typically treated as a statistical nuisance and absorbed through spatial lags, clustered errors, or lagged dependent variables, rather than estimated and examined as a meaningful manifestation of the migration process.

An alternative, more recent approach is to employ Social Network Analysis (SNA) to model macroscale migration networks: systems of aggregate flows between locations (the *nodes*) connected by the flows of people moving between them (the *directed*,

often *weighted ties*; Danchev & Porter, 2021).² The network-analytic approach treats the full configuration of bilateral flows as the unit of analysis and asks whether aggregate properties of this configuration influence observed flow patterns. Several recent studies have advanced this agenda, showing that properties such as transitivity, reciprocity, and hierarchical structures may organise both internal and international migration flows (Huang & Butts, 2024, 2023; Leal, 2021; Leal & Harder, 2021, 2023; McMillan, 2024; Windzio, 2018). These studies demonstrate that considering network structures and interdependence improves both substantive interpretation and model fit. However, they operate at the systemic level: interdependence is estimated as an aggregate property of the migration network itself, rather than as a feature of individual destination choices.

This study's analytical strategy shares foundations with previous SNA applications but takes a different position. The proposed model retains a distinctive individual-level interpretation: each migrant's destination choice is modelled as a function of destination attributes, bilateral characteristics, and the choices of others.

5.3 Research Questions and Analytical Strategy

So far, this chapter has argued that relational structures generate interdependence between migration decisions and that standard macro-models may implicitly assume the opposite: that migration decisions are conditionally independent once contextual factors are taken into account. If, however, interdependence operates through relational structures that are correlated with macro-level drivers, models assuming independence may yield biased estimates of other migration predictors.

This claim requires empirical scrutiny. If interdependence exerts only a marginal influence on migration flows or operates orthogonally to commonly used predict-

²The notion of macroscale migration networks is distinct from that of 'microscale' networks of migrants or 'migrant networks', which, instead, focus on migrant communities, their interpersonal ties, and their interactions (Danchev & Porter, 2021).

ors, treating migration decisions and flows as independent would be a reasonable approximation. Assessing the empirical relevance of interdependence is therefore not only a theoretical concern, but a practical one. This constitutes the central aim of the present chapter, where I ask: (1) *To what extent does interdependence structure migration flows beyond conventional predictors?* (2) *Does ignoring interdependence distort estimates of those predictors?*

To address these questions, I employ ELMos, a class of statistical network models designed to study interdependence in mobility and migration networks (Block et al., 2022). ELMos can be understood as multinomial logit models extended with endogenous parameters that capture interdependence between individual migration decisions. The outcome of interest is each individual's destination choice, modelled as a function of destination attributes, bilateral characteristics linking origins and destinations, and endogenous parameters representing origin and dyadic interdependence and capturing the behaviour of other migrants in the system. The total flows between countries are then understood as an aggregation of individual migration decisions. By construction, the model conditions on each individual's origin, so inference concerns destination selection given the decision to migrate. When the endogenous parameters are set to zero, the model reduces to a standard multinomial logit. The model's capacity to represent interdependence directly is inspired by existing statistical network models, most notably Exponential Random Graph Models (ERGMs; Lusher et al., 2012), but adapted to the specific structure of migration data.

Unlike previous statistical network analyses of migration, which typically take location, dyads, or the entire network as the fundamental unit of analysis, the ELMO operates at the individual level while preserving the network structure, accommodating both the direction and magnitude of movement between locations. Origin and dyadic interdependence are introduced as estimable parameters to assess their magnitude, direction, and consequences for conventional predictors. Each of the two forms of interdependence is operationalised as one parameter capturing opposite ends of a

signed continuum, with origin interdependence ranging from concentration (positive) to evenness (negative), and dyadic interdependence from reciprocity (positive) to asymmetry (negative).

5.3.1 Inference to Relational Structures

The endogenous parameters estimated by the ELMo capture statistical regularities in the migration system: tendencies for migration decisions to be linked to one another beyond what conventional predictors included in the model may explain. These regularities are consistent with relational structures, but they do not by themselves identify the specific mechanisms that generate them. Origin interdependence may arise from peer influence, organisational sourcing pipelines, or migrant competition; dyadic interdependence may reflect bilateral exchange norms, institutional strategies, or their absence. Identifying particular mechanisms requires additional information on the relational structures involved in generating interdependence. The claim here is different: that interdependence, irrespective of its precise origins, is an empirical trace of underlying migration mechanisms, and therefore a fundamental ingredient in understanding the spatial organisation of movement.

To assess this claim and the adequacy of the ELMo, I pursue four complementary strategies. First, in the empirical application, I specify models that include extensive bilateral predictors alongside origin and destination fixed effects. This ensures that the interdependence parameters capture relations in the data not attributable to observed or time-invariant characteristics of origins, destinations, or corridors. While no model specification can claim completeness, in the application I include the country-level drivers most widely used in the ISM literature. Second, I estimate nested models. If interdependence parameters merely absorb omitted variable bias, their magnitude or significance should diminish once additional predictors are introduced. This directly addresses the first research question. Third, to address the second research question, I examine whether modelling interdependence reshapes conventional pre-

dictors. If independence-based models are incomplete but unbiased, estimates should remain stable. If they are biased by omitted interdependence, coefficients should shift systematically once the endogenous parameters are included. Fourth, I use a comparative design and estimate the models separately for degree and Erasmus mobility: two systems involving the same countries in the same period but organised through different relational structures. If interdependence parameters reflect artefacts of unobserved country-level, bilateral characteristics, they should behave similarly across both systems. If they capture genuine differences in relational structures, they should diverge in theoretically expected ways.

Crucially, none of these diagnostics constitutes a definitive causal test. The logic is abductive: if theoretically expected patterns emerge consistently across these tests, under a design that controls for plausible confounders, the interpretation that the endogenous parameters capture genuine interdependence becomes more compelling than alternative explanations. The claim is not that the underlying relational processes have been directly identified, but that interdependence is *explanatorily relevant*: modelling it changes conclusions about conventional predictors in ways that assuming independence does not. This approach is appropriate when the mechanisms generating interdependence cannot be observed directly, and it aligns with the broader epistemological position that what matters for explanation is counterfactual dependence between *explanans* and *explanandum*, not complete specification of the causal chain (Ylikoski, 2011).

Applying the model to ISM, a setting where relational structures are comparatively well understood, provides a test of whether the ELMo recovers theoretically expected forms of interdependence. Success in this case would strengthen confidence in extending the approach to migration contexts where the underlying relational structures are less visible or less thoroughly theorised.

5.4 Empirical Case: European Student Mobility

I apply the outlined general framework to a particular, well-studied migration case: the international movement for higher education purposes, or International Student Mobility (ISM). ISM is among the largest and most rapidly growing forms of high-skilled migration, both in terms of *degree mobility*, where students move abroad to complete an entire degree programme, and *credit mobility*, where students spend a limited period abroad (usually less than one year) while remaining enrolled at their home institution (Choudaha, 2017; de Wit et al., 2012; Gümüş et al., 2020). Within the ISM landscape, Europe occupies a distinctive position: in the 1990s, the EU's internationalisation agenda saw the initiation of the Bologna Process, which standardised higher education curricula, degrees, and enrolment requirements within the region, considerably facilitating degree mobility (Recchi, 2015; Teichler, 2019, 2012). At the same time, the EU is home to the Erasmus programme, the largest and most important example of credit mobility worldwide, launched in 1987 to promote academic exchange, cultural interaction, and intra-European integration (Recchi, 2015). Since then, the Erasmus programme has grown to involve thousands of HEIs and hundreds of thousands of students each year, under the support of EU funding and formal exchange agreements (European Commission, 2017).

Across both degree and Erasmus mobility, prior research has identified a broad set of relevant migration drivers. Geographic distance between origin and destination countries has been found to be negatively associated with degree mobility, while it appears to play a more limited role in Erasmus mobility, reflecting the costs and frictions associated with longer-term moves (Abbott & Silles, 2016; Beerkens et al., 2016; Beine et al., 2014; Perkins & Neumayer, 2014; Souto-Otero et al., 2013; Van Bouwel & Veugelers, 2013; Vögtle & Windzio, 2022). Economic differentials between locations have been found to matter: student flows tend to be larger when moving from lower-income origins to higher-income destinations, and greater dissimilarity in GDP per capita is

associated with higher levels of both degree and Erasmus mobility (Abbott & Silles, 2016; Beine et al., 2014; Perkins & Neumayer, 2014; Weber & Van Mol, 2023). Similarly, past (non-student) migration from origin to destination countries has been linked to higher student mobility in both forms (Beine et al., 2014; Murphy-Lejeune, 2002; Van Mol & Timmerman, 2014). Similarity in language between origin and destination shows more heterogeneous effects: for degree mobility, sharing an official language has been found to exert a weakly positive or null effect (Barnett et al., 2016; Ovchinnikova et al., 2023; Perkins & Neumayer, 2014; Van Bouwel & Veugelers, 2013); for Erasmus mobility, language differences have been found to have a negative effect, with students frequently seeking to acquire or improve foreign language skills during short exchanges abroad (Lesjak et al., 2015; Perez-Encinas et al., 2020; Rodríguez González et al., 2011; Van Mol, 2014).

Alongside these drivers, research on ISM has also investigated whether student destination choices are interdependent. Qualitative and small-scale studies suggest that origin interdependence is dominant in degree mobility: students aiming to pursue a degree abroad rely on the experiences and advice of co-national peers, family members, and older students, and often follow the exact same trajectories (Beech, 2015; Collins, 2008; Jayadeva, 2024, 2020; Perez-Encinas et al., 2020; Van Mol & Timmerman, 2014).

By contrast, Erasmus mobility is expected to be shaped primarily by dyadic interdependence. Erasmus mobility is built on bilateral inter-institutional agreements so that HEIs directly agree on exchanging students with each other. Often, there is the explicit and formal requirement that participating HEIs will send and receive students in balanced numbers (eQuATIC Project & European University Foundation, 2019). However, institution-specific evidence indicates that HEIs may pursue asymmetric objectives when entering partnerships, including satisfying their students' demand for attractive destinations, strengthening their international profile, or securing financial or symbolic prestige (Ferreira-Pereira & Mourato Pinto, 2021; Maringe & de Wit, 2016). Under both scenarios, HEIs retain decisive authority over which destinations are

available to their students as part of Erasmus exchanges; individual preferences, even when influenced by migration drivers, are exercised within institutionally defined choice sets (Beerkens et al., 2016; Teichler, 2004).³

The active structuring of Erasmus flows by HEIs has no direct analogue in degree mobility, where students' destination choices are not restricted to predefined partner sets. This contrast makes European ISM a particularly strong empirical test of the proposed analytical strategy. Both mobility types involve similar populations (young European adults moving abroad for higher education) and are subject to comparable contextual conditions and macro-level drivers. What differs is the relational structure organising each mobility type: individual choice and interpersonal networks in degree mobility; bilateral institutional agreements in Erasmus. If the ELMo successfully captures interdependence rather than unobserved confounding, the origin and dyadic parameters should differ across ISM types. More specifically, one would expect stronger positive origin interdependence in degree mobility, where peer and social influence operate with minimal institutional constraints, and stronger dyadic interdependence in Erasmus mobility, where the direction of the corresponding parameter (positive *vs* negative) sheds light on the underlying institutional logic (balanced exchange *vs* strategic asymmetry).

5.5 Data

The outcome variable in the ELMos is the decision to migrate abroad to pursue a higher education degree (degree mobility) or as part of an Erasmus exchange (credit

³In the Erasmus case, another potential source of origin interdependence might be migrant competition. Students who intend to participate in the Erasmus programme may be selected by their home institution, typically on the basis of academic performance, language proficiency, and motivation to study abroad (Teichler, 2004). Because the number of available exchange spots at specific destination institutions may be limited, competition would intensify when many students target the same HEI. Limited qualitative research on Erasmus participants suggests that most students primarily seek the opportunity to go abroad, rather than access to a specific destination: under such conditions, students may strategically adjust their preferences in response to anticipated competition, redirecting applications toward less popular HEIs to increase their chances of being selected (Krzaklewska, 2008; Murphy-Lejeune, 2002).

mobility). I focus on all moves that occurred within 29 European countries⁴ in the calendar year 2019. I select 2019 as it was the last year in which the UK, one of the most important ISM destination countries, was part of the EU and participated as a full member of the Erasmus programme, and because it is the most recent period before the COVID-19 pandemic, which substantially disrupted ISM and higher education systems worldwide, with potential long-lasting consequences (Mihut et al., 2025; Yang & Yi, 2023).

Data on degree mobility are produced jointly by the Institute for Statistics (UIS) of the United Nations Educational, Scientific, and Cultural Organisation (UNESCO), the Organisation for Economic Co-operation and Development (OECD), and Eurostat (UIS, 2023). Information on foreign students' enrolment is collected by individual HEIs, harmonised by the three agencies, and disseminated in aggregate form, so that degree mobility is observed as country-to-country aggregate flows for each calendar year. Data on Erasmus mobility come from the European Commission (2021), which compiles individual-level, anonymised administrative registers of all student moves conducted under the Erasmus programme. These registers report detailed information on each mobility episode at the individual level, including the student's nationality, course of study, sending and receiving institution, and the type and duration of the exchange.

The administrative and register-based nature of both data sources is particularly important to evaluate the role of interdependence in migration processes. Unlike most data sets of international migration flows, which typically rely on model-based reconstructions from migrant stocks, the degree and Erasmus flows constitute observed registrations and not estimated counts. This reduces the risk that any observed interdependence in flows stems from artefacts of flow estimation, rather than a genuine

⁴The selected countries are, in alphabetical order: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Spain, Sweden, and the UK.

migration feature (see eg Daňko et al., 2024, and Yildiz and Abel, 2021, for relevant discussions of the quality of migration flow data).

To ensure comparability between the two ISM types, I aggregate individual Erasmus moves by country of origin and destination. I further restrict the Erasmus sample to students who are nationals of the country where their university is located—mirroring the definition of degree mobility based on students’ (foreign) nationality—and to those who completed an academic exchange.⁵

These restrictions yield populations of 222,773 degree-mobile students and 182,106 Erasmus-mobile students. Given computational constraints, I draw a random sample of 20,000 students (approximately 10%) from each population for the statistical models. Sampling was stratified by origin–destination pairs to ensure that the distribution of students across corridors matches and is representative of the full data. A description of the full and sampled populations is included in Table 5.1.

To account for established ISM drivers, I include a set of bilateral predictors, capturing ties between origin and destination countries. These include: geographic proximity, measured through shared borders (contiguity) and logged great-circle distance between country capitals; lagged economic and demographic similarity, operationalised as similarity in GDP per capita and in population size in 2018; the presence of a shared official language; lagged total migration flows (of all types) from origin to destination in 2018; and shared status as an Erasmus founding-member country.⁶ Information on these drivers was sourced from the Centre d’Etudes Prospectives et d’Informations Internationales (CEPII) Gravity database (Conte et al., 2021), the World Bank (2025a, 2025b), and the Human Migration Database (HMigD; Daňko, 2023; Daňko et al., 2024), as described in Appendix C.1.

⁵I therefore exclude cases where Erasmus mobility was pursued for traineeships, short-term language courses, volunteering, and other non-study exchanges.

⁶The Erasmus programme was initially launched in eleven countries (Belgium, Denmark, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain, and the UK) and saw progressive expansions in 1992, 1995, 2004, 2007, 2013, 2014, and 2019, largely mirroring the EU’s enlargement waves (Recchi, 2015). I include this predictor to account for potential path dependence and network segmentation around founding and non-founding countries in the sample.

Table 5.1: Summary of full and sampled student populations. Country-specific aggregate outflows (out) and inflows (in) for degree (D) and Erasmus (E) mobility.

Country name	ISO2 code	D-out (total)	D-out (sample)	D-in (total)	D-in (sample)	E-out (total)	E-out (sample)	E-in (total)	E-in (sample)
Austria	AT	4,581	412	9,434	848	3,168	348	4,443	489
Belgium	BE	6,657	596	26,367	2,367	5,497	604	5,986	657
Bulgaria	BG	7,228	648	937	82	742	82	752	84
Cyprus	CY	5,032	451	732	65	184	20	532	60
Czechia	CZ	3,399	305	5,885	530	3,937	433	6,024	663
Germany	DE	41,561	3,732	14,128	1,269	28,741	3,156	15,560	1,712
Denmark	DK	1,535	136	8,611	774	1,920	210	3,437	380
Estonia	EE	1,112	99	474	40	451	50	1,186	129
Spain	ES	11,296	1,017	6,252	562	31,586	3,470	29,378	3,226
Finland	FI	3,094	277	982	87	3,413	375	6,207	682
France	FR	29,795	2,677	11,792	1,060	25,962	2,852	17,213	1,889
Greece	GR	11,826	1,060	984	87	3,036	333	2,266	250
Croatia	HR	1,482	133	75	7	1,172	127	1,438	155
Hungary	HU	3,492	311	1,912	171	2,491	275	3,620	396
Ireland	IE	5,295	473	2,185	197	2,361	256	4,581	500
Iceland	IS	824	74	185	16	199	23	447	49
Italy	IT	21,994	1,976	3,151	285	28,127	3,089	18,368	2,019
Lithuania	LT	3,461	312	270	22	1,386	152	1,864	204
Luxembourg	LU	4,402	394	593	52	281	30	99	9
Latvia	LV	1,589	141	372	34	745	82	1,152	128
Malta	MT	470	41	200	18	286	32	328	35
Netherlands	NL	8,197	738	61,235	5,500	7,257	798	8,336	920
Norway	NO	4,935	445	1,156	103	2,272	248	5,551	608
Poland	PL	7,161	643	1,514	136	7,630	839	9,688	1,064
Portugal	PT	3,970	358	1,331	123	5,836	640	9,846	1,082
Romania	RO	9,639	867	1,328	119	2,932	321	1,906	210
Sweden	SE	3,522	317	2,872	255	2,602	285	6,862	753
Slovak Republic	SK	8,020	719	1,584	143	1,677	184	1,283	140
United Kingdom	UK	7,203	648	56,231	5,048	6,215	686	13,753	1,507

5.6 Methods

5.6.1 Formulation of the Network Model

The two ISM types are analysed using the ELMo, as described in Hoffman and Block (2026). Mathematically, the ELMo defines the probability that an individual i selects a destination j_1 from a set of j potential locations, conditional on the migration behaviour of all other individuals in the network, denoted as x^{-i} . This probability is given by:

$$p(i \rightarrow j_1 \mid \theta, x^{-i}) = \frac{\exp(\sum_k \theta_k t_k(i \rightarrow j_1 \mid x^{-i}))}{\sum_{m=1}^M \exp(\sum_k \theta_k t_k(i \rightarrow j_m \mid x^{-i}))}, \quad (5.1)$$

where θ is a vector of model parameters, m indexes destinations, and $t(\cdot)$ are statistics describing the network and serving as predictors.

Model parameters θ_k can be interpreted analogously to coefficients in other log-linear models (eg logistic regression) as guiding the likely destination of individuals. Positive (negative) coefficients indicate that individuals tend to move to (avoid) destinations that score higher on the respective statistic t , which encodes, for example, whether origin and destination speak the same language, or how many other migrants move from i 's origin to destination j_1 . The exact size of parameters can be interpreted identically to classical multinomial logit coefficients. When $t(\cdot)$ includes only driver-based predictors (eg country-level characteristics) and is orthogonal to x^{-i} , the ELMo reduces to a standard multinomial logit model. In the proposed specification, however, the statistics $t(\cdot)$ are allowed to include the movement of other individuals, enabling the model to represent interdependence directly. In intuitive terms, the formula captures how likely individual i is to choose destination j_1 given what everyone else in the network is doing.

Based on the conditional formulation in Equation (5.1), it is possible to define the joint probability of observing the entire migration network as:

$$p(X = x) = \frac{\exp(\sum_k \theta_k s_k(x))}{\sum_{x' \in \mathbb{C}} \exp(\sum_k \theta_k s_k(x'))}, \quad (5.2)$$

where X is a random migration network (akin to a random variable), x is one (usually the observed) realisation, $s_k(x)$ are sufficient statistics of the network, and $\mathbb{C}$ is the set of all possible network realisations. The link between the conditional and joint formulations is given by:

$$t(i \rightarrow j_1 | x^{-i}) = s(i \rightarrow j_1, x^{-i}) - s(x^{-i}), \quad (5.3)$$

that is, the predictor at the individual level t corresponds to the change in the network statistic s when adding the move of individual i , $i \rightarrow j_1$, to the otherwise complete network x^{-i} . Thus, the sum of all individual contributions of t equals s .

This joint probability is the basis for estimation of model parameters. However, the normalising constant in Equation (5.2) sums over all possible configurations $\mathbb{C}$ the network can take. For networks of realistic size, this is computationally intractable: with N individuals moving within M destinations, the cardinality of $\mathbb{C}$ is M^N . In the empirical application, where I consider two migration networks (one for degree mobility, one for Erasmus) of 20,000 individuals and 29 countries each, the cardinality is $|\mathbb{C}| = 29^{20,000}$. Estimation thus relies on specialised Markov chain Monte Carlo (MCMC) techniques, as originally proposed by Snijders (2002) for statistical network models and adapted to mobility network models by Block et al. (2022).⁷

5.6.2 Formulation of the Interdependence Parameters

The two interdependence forms, origin and dyadic, enter the model in Equation (5.1) as signed parameters. Origin interdependence is operationalised as:

⁷It should be noted that, even with MCMC, computation time is a function of the data size: the estimated models run for several days on high-performance computing clusters using the statistical software R (version 4.3.2; R Core Team, 2023).

$$t_{\text{origin}}(i \rightarrow j_1 \mid x^{-i}) = g_{\text{mar}}(x_{o_i, j_1}^{-i}), \quad (5.4)$$

where

$$g_{\text{mar}}(y) = -1 + \sum_{k=0}^y e^{-\log(\alpha)k}, \quad (5.5)$$

and o_i denotes the origin of individual i . The value of t_{origin} equals zero if nobody else moves in the same origin–destination corridor, $o_i \rightarrow j_1$, and increases with the number of other individuals moving along the corridor, approaching $\frac{1}{\alpha-1}$ as flows along that corridor grow large. The scaling parameter α governs the rate at which this asymptote is reached and the maximum value of t_{origin} .⁸ A positive estimate of θ_{origin} would indicate concentration dynamics, whereby individuals are more likely to select destinations that are already popular among others from the same origin. A negative estimate would indicate the opposite tendency, whereby individuals systematically avoid crowded corridors (evenness).

Dyadic interdependence is operationalised as:

$$t_{\text{dyadic}}(i \rightarrow j_1 \mid x^{-i}) = g_{\text{mar}}(x_{j_1, o_i}^{-i}), \quad (5.6)$$

mirroring the origin interdependence term, but reversing the direction of the corridor to capture the extent to which mobility from o_i to j_1 is associated with flows in the opposite direction, $j_1 \rightarrow o_i$. A positive estimate of θ_{dyadic} would reflect symmetric and reciprocal flows between two locations (reciprocity). A negative estimate would indicate that migration predominantly runs in one direction (asymmetry).

The mathematical formulation of both interdependence statistics is guided by substantive and technical considerations. The statistics are constructed to mitigate well-known issues of model degeneracy while yielding parameters with a clear

⁸For both interdependence parameters, I set $\alpha = 1.05$ as a grid search in steps of 0.05 indicated best fit at this parametrisation. This corresponds to a maximum value of 20.

interpretation, following Hoffman and Block (2026) and adapting principles originally developed for ERGMs (Snijders et al., 2006).

5.6.3 Model Specification

The empirical analysis focuses on a single cross-section (the year 2019). This choice reflects a substantive interest in the structural organisation of ISM and the interdependence embedded in contemporaneous flows, rather than in temporal dynamics. Because each year involves a different cohort of students, modelling longitudinal change at the individual level would not be informative in this context. Moreover, incorporating lagged student flows as predictors would introduce autoregressive dependencies that obscure the interpretation of the endogenous parameters intended to capture interdependence (Block et al., 2018). Restricting attention to a single year allows me to isolate the structural features of the migration networks and model the interdependence parameters that are central to the inquiry.

To answer the research questions, I estimate four models for each ISM type. The first model acts as a baseline and only includes destination fixed effects, as the modelling strategy already conditions on the origin marginal by taking individual origins as given.⁹ The four models are defined as follows:

Model 1 = Baseline model with origin and destination conditioning

Model 2 = Model 1 + bilateral predictors (standard predictors of ISM)

Model 3 = Model 1 + interdependence parameters (origin and dyadic)

Model 4 = Model 2 + interdependence parameters (full model)

This modelling strategy makes it possible to assess whether interdependence improves model fit and whether it reshapes inference about conventional drivers. In line with

⁹In other words, all models fully control for the number of incoming and outgoing students per country.

the theoretical framework, the bilateral predictors capture known drivers of ISM flows, while the origin and dyadic interdependence parameters capture relational tendencies. Because all specifications include origin conditioning and destination fixed effects, the models account for all variance related to country-specific (monadic) predictors. Country-level information is included in dyadic measures of similarity. For any country predictor V and a dyad composed of countries i and j , similarity is defined as $\text{sim}_{ij}^V = 1 - \frac{|v_i - v_j|}{\max(V) - \min(V)}$, where v_i and v_j denote the values of V for countries i and j , respectively. This yields a symmetric, bounded measure ranging from zero (maximum dissimilarity) to one (perfect similarity).

5.6.4 Goodness-of-Fit

To evaluate model performance, I adapt standard measures of fit from log-linear modelling frameworks, in particular the (pseudo-)log-likelihood, the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC). The (pseudo-)log-likelihood is computed by summing the log-transformed predicted probabilities assigned to each individual's actually chosen destination:

$$\log \mathcal{L} = \sum_{i=1}^N \log \hat{p}_{ij_i}, \quad (5.7)$$

where $\hat{p}_{ij_i}$ is the model's predicted probability that individual i selects their observed destination j , conditional on the observed state of all other individuals as given in Equation (5.1). When interdependence is included, this becomes a pseudo-log-likelihood, because individual probabilities are artificially factorised, which approximates the actual model likelihood given in Equation (5.2). In the absence of such terms, Equation (5.7) is simply the log-likelihood. The AIC and BIC are derived from the (pseudo-)log-likelihood in conventional ways.

5.6.5 Relative Importance of Predictors

Given well-known difficulties in interpreting parameters in log-linear models (see eg Mood, 2010), I introduce an additional metric to understand the extent to which different predictors may influence modelled outcomes. I assess the relative importance (RI) of each predictor t in the model by quantifying its marginal influence on predicted probabilities. This measures how much the model's predictions change when a single predictor is excluded from the estimation, achieved by artificially setting the associated parameter θ to zero while keeping all other parameters fixed at their estimated values. This approach mirrors the computation of marginal effects in other regression models and draws on techniques developed for Stochastic Actor-Oriented Models (SAOMs) by Indlekofer and Brandes (2013).

Formally, I use simulations to generate counterfactual versions of the model in which each predictor is sequentially removed or 'knocked out' by setting its associated parameter to zero. For each modified model, I recompute the predicted probability distribution across destinations for each individual and compare these predictions to those from the full model by calculating the absolute differences in predicted probabilities and summing them across all individuals and outcomes. The raw influence score of a given parameter is then the total amount by which predictions shift due to the parameter's removal. Raw influence scores are normalised to sum to one, yielding a relative importance score for each predictor:

$$\text{Relative Importance}_k = \frac{\sum_{i=1}^N \sum_{j=1}^M \left| \hat{p}_{ij}^{\text{full}} - \hat{p}_{ij}^{(-k)} \right|}{\sum_{k'=1}^K \sum_{i=1}^N \sum_{j=1}^M \left| \hat{p}_{ij}^{\text{full}} - \hat{p}_{ij}^{(-k')} \right|}, \quad (5.8)$$

where $\hat{p}_{ij}^{\text{full}}$ is the predicted probability from the full model that an individual i chooses destination j , and $\hat{p}_{ij}^{(-k)}$ is the corresponding prediction from the model when the predictor k is removed. Applied to the interdependence parameters, this diagnostic speaks directly to the first research question: if origin and dyadic

interdependence account for a negligible share of explanatory power, the theoretical emphasis on interdependence may be empirically unwarranted; if they account for a substantial share, models that assume independence would be missing a primary organising principle.

5.6.6 Corridor-Level Model Diagnostics

Because ELMos yield full probability distributions over all potential destinations for each individual, it is possible to evaluate model performance at the level of migration corridors rather than only through aggregate fit statistics. I implement two complementary diagnostic strategies.

First, I assess corridor-level prediction bias directly from the predicted probabilities. For each origin–destination pair, I sum the predicted probabilities across all individuals and compare the resulting predicted share to the observed share of total migration. This results in a relative bias metric indicating the extent to which each corridor is over- or under-predicted. Comparing bias within and between model specifications reveals which corridors are more difficult to predict.

Second, I use the predicted probabilities as a generative device to simulate synthetic migration networks. For each individual, I draw a destination from the predicted probability distribution; aggregating these draws yields a simulated flow matrix. Repeating this process produces a distribution of plausible networks under each model specification. Comparing these simulated networks to the observed ones provides a further diagnostic, allowing me to examine not only whether a model's mean predictions are accurate, but also whether the observed flows fall within the variability range implied by the model.

5.7 Results

A description of the degree and Erasmus flow samples is shown in Figure 5.1. In each panel, rows represent the origins and columns the destinations, with cell shading reflecting the size of bilateral flows on a log-scaled gradient. From this figure, degree mobility displays visibly stronger concentration patterns, with a small number of destinations—notably the Netherlands and the UK—receiving disproportionately large inflows from many origins. Several countries in the sample, such as Cyprus, Croatia, Greece, and Latvia, exhibit very small incoming degree mobility flows. Conversely, Erasmus exchanges display a more diffuse and even spatial distribution, with moderate flows distributed across a larger number of corridors.

5.7.1 Model Results

Estimates of the statistical network models for degree and Erasmus mobility are reported in Table 5.2. MCMC diagnostics showing model convergence in Models 3–4 are reported in Appendix C.2, Figures C.1 to C.4. For both ISM types, the goodness-of-fit statistics show that the full model (Model 4) has the best fit. These improvements in fit, compared in particular to Model 2, confirm that incorporating interdependence effects enhances predictive accuracy while retaining model parsimony. I complement model estimates with the relative influence of the Model 4 parameters in Figure 5.2.

From Table 5.2, I find that across all network specifications (Models 3–4) the interdependence parameters are statistically significant. This underscores their role in capturing relations in migration decisions that are not easily explained by independent responses to origin–destination characteristics. The origin interdependence parameter is positive and statistically significant for both forms of ISM (degree: 0.109, $p < 0.001$; Erasmus: 0.041, $p < 0.001$ in Model 4), indicating that flows from the same origin tend to concentrate in a subset of destinations rather than being evenly distributed across all available migration corridors.

International Student Mobility Flows in Europe (2019)

Darker cells denote larger flows. Blank cells denote zero flows.

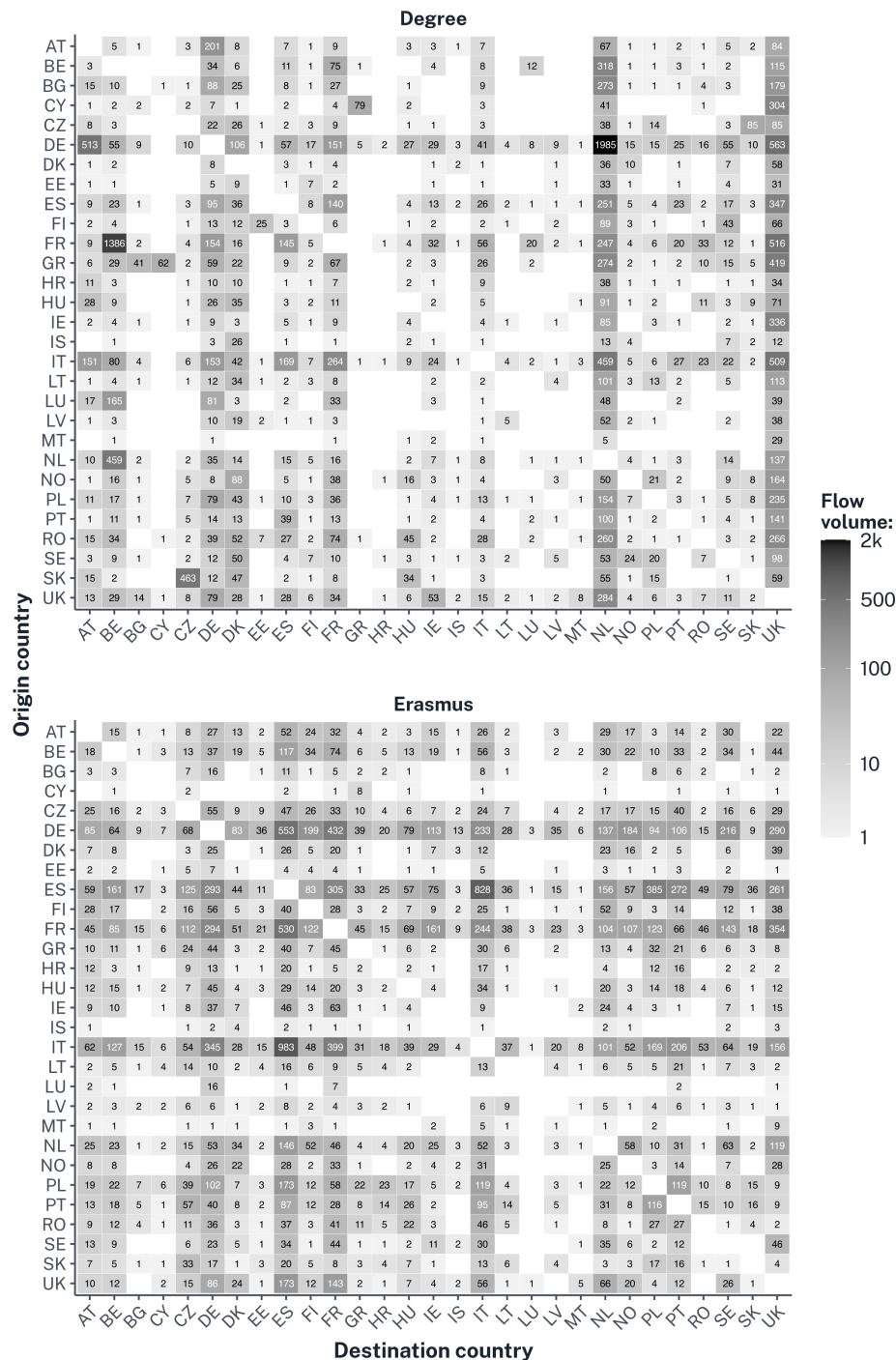


Figure 5.1: Bilateral flows of degree- and Erasmus-mobile students within 29 European countries in the year 2019. Random samples of 20,000 individuals per mobility type were drawn from the total populations using stratified sampling by origin–destination country dyads, to ensure consistency in the corridor-level distribution of students. Each cell reports the number of students that moved from the origin (row) to the destination (column).

Put differently, this shows that students from the same origin are more likely to go to the same destination(s). The origin interdependence coefficient is larger for degree than Erasmus mobility; moreover, from the RI plots in Figure 5.2, it is possible to see that in degree mobility, origin interdependence accounts for 40% of relative explanatory importance, exceeding any conventional bilateral predictor. In Erasmus mobility, origin interdependence is weaker, but still ranks among the most influential predictors (RI = 18%), indicating that concentration dynamics may operate even under institutional constraints.

The dyadic interdependence parameter shows a sharper divergence between the two ISM types. For degree mobility, the coefficient is positive and statistically significant (0.003, $p < 0.001$ in Model 4), suggesting a modest tendency for flows to occur in both directions in a country pair. In other words, a degree-mobile student is more likely to select a destination that experiences outflows towards that student's country of origin, holding all other predictors constant. However, the RI shows that dyadic interdependence is a rather unimportant predictor of degree mobility (RI < 3%): foreshadowing the discussion, this pattern might reflect unmeasured origin–destination similarities rather than genuine reciprocal exchange mechanisms, as degree mobility lacks an explicitly bilateral institutional architecture.

Conversely, the negative coefficient in Erasmus mobility indicates systematic asymmetry: flows from one country to another are associated with lower flows in the opposite direction, net of overall sending and receiving volumes (-0.011 , $p < 0.001$ in Model 4). In other words, there is less reciprocity in exchanges than non-systematic, 'random' migration would predict, indicating that, between pairs of countries in the Erasmus network, one predominantly acts as the sender and the other as the receiver of mobility. Further, the RI analysis shows that dyadic interdependence is a relevant Erasmus predictor (RI = 8%), albeit less important than origin interdependence.

Table 5.2: Predicting degree and Erasmus student mobility flows using Endogenous Log-linear Models (ELMos).

	Degree				Erasmus			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
<i>Bilateral predictors:</i>								
Contiguity		0.749*** (0.031)		0.529*** (0.027)		-0.186*** (0.029)		-0.121*** (0.028)
Distance		0.211*** (0.022)		0.071*** (0.020)		0.125*** (0.014)		-0.114*** (0.020)
GDP per capita		0.566*** (0.100)		0.770*** (0.079)		2.626*** (0.080)		1.560*** (0.097)
Population size		-0.259*** (0.056)		-0.090 [†] (0.049)		-0.219*** (0.048)		-0.169*** (0.046)
Language		1.244*** (0.039)		0.931*** (0.033)		-0.316*** (0.057)		-0.273*** (0.051)
General migration		0.392*** (0.012)		0.215*** (0.011)		0.092*** (0.008)		-0.044*** (0.011)
Erasmus founding		0.257*** (0.048)		0.397*** (0.039)		-0.248*** (0.037)		-0.327*** (0.036)
<i>Interdependence predictors:</i>								
Origin			0.140*** (0.002)	0.109*** (0.002)			0.048*** (0.003)	0.041*** (0.003)
Dyadic			0.007*** (0.0005)	0.003*** (0.0005)			-0.008*** (0.001)	-0.011*** (0.001)
<i>Conditioning (fixed effects):</i>								
Origin	✓	✓	✓	✓	✓	✓	✓	✓
Destination	✓	✓	✓	✓	✓	✓	✓	✓
<i>Fit statistics:</i>								
Log-likelihood	-45,046	-40,272	-40,462	-38,523	-57,434	-55,426	-55,083	-54,852
AIC	90,148	80,614	80,984	77,119	114,924	110,923	110,225	109,779
BIC	90,369	80,890	81,222	77,412	115,146	111,200	110,462	110,071
Observations:	20,000	20,000	20,000	20,000	20,000	20,000	20,000	20,000

Notes: Standard errors in parentheses. Significance codes reported as: [†]: 0.1, *: 0.05, **: 0.01, ***: 0.001.

All models condition on origin and destination. Model 1 is the baseline. Model 2 is a traditional macro-model with bilateral predictors of country-to-country migration flows. Model 3 includes only interdependence effects. Model 4 is the full model.

With respect to the bilateral predictors, results align broadly with previous findings on the drivers of ISM. Following the RI plots, the most important drivers for degree-mobile students are contiguity, speaking the same language, and studying in countries that are also popular targets for other co-national migrants, conditional on origin and

destination characteristics. GDP similarity between origin and destination appears to play a positive but relatively minor role. For Erasmus mobility, the two most important predictors in terms of RI are GDP similarity and Erasmus founding membership. However, joint status as Erasmus founding members is negative for Erasmus mobility ($-0.327, p < 0.001$ in Model 4). Simply put, Erasmus exchanges are more likely to occur between founding and non-founding member countries than within these categories. The presence of a common language is negatively associated with Erasmus mobility ($-0.273, p < 0.001$ in Model 4), consistent with findings that Erasmus students often choose destinations where they can develop foreign language skills; however, in terms of RI, this driver appears to be relatively unimportant ($RI < 3\%$).

Additionally, most coefficient estimates change substantially between Model 2 and Model 4; that is, after the inclusion of interdependence parameters. For example, the 'Distance' predictor is positive and significant in both mobility types, but decreases in magnitude or reverses sign after inclusion of the interdependence parameters. This means that, under Model 2 and controlling for contiguity, student mobility would occur between countries that are further apart. Yet, once the interdependence of migration is accounted for, distance is no longer related to destination selection in the same way, particularly for Erasmus mobility. This reversal indicates that the apparent positive association between distance and Erasmus flows in independence-based models reflects the presence of highly popular and concentrated long-distance corridors. Once interdependence is modelled explicitly, students exhibit no additional propensity to choose countries that are geographically distant from their origin. Similar patterns of sign reversal and magnitude change appear for other bilateral predictors, such as population size in degree mobility and general migration in Erasmus mobility, indicating that models assuming independence may systematically misattribute explanatory weight when interdependence structures are present.

Relative Importance of Model 4 Predictors

Change in predicted probabilities when the effect of a predictor is set to zero.

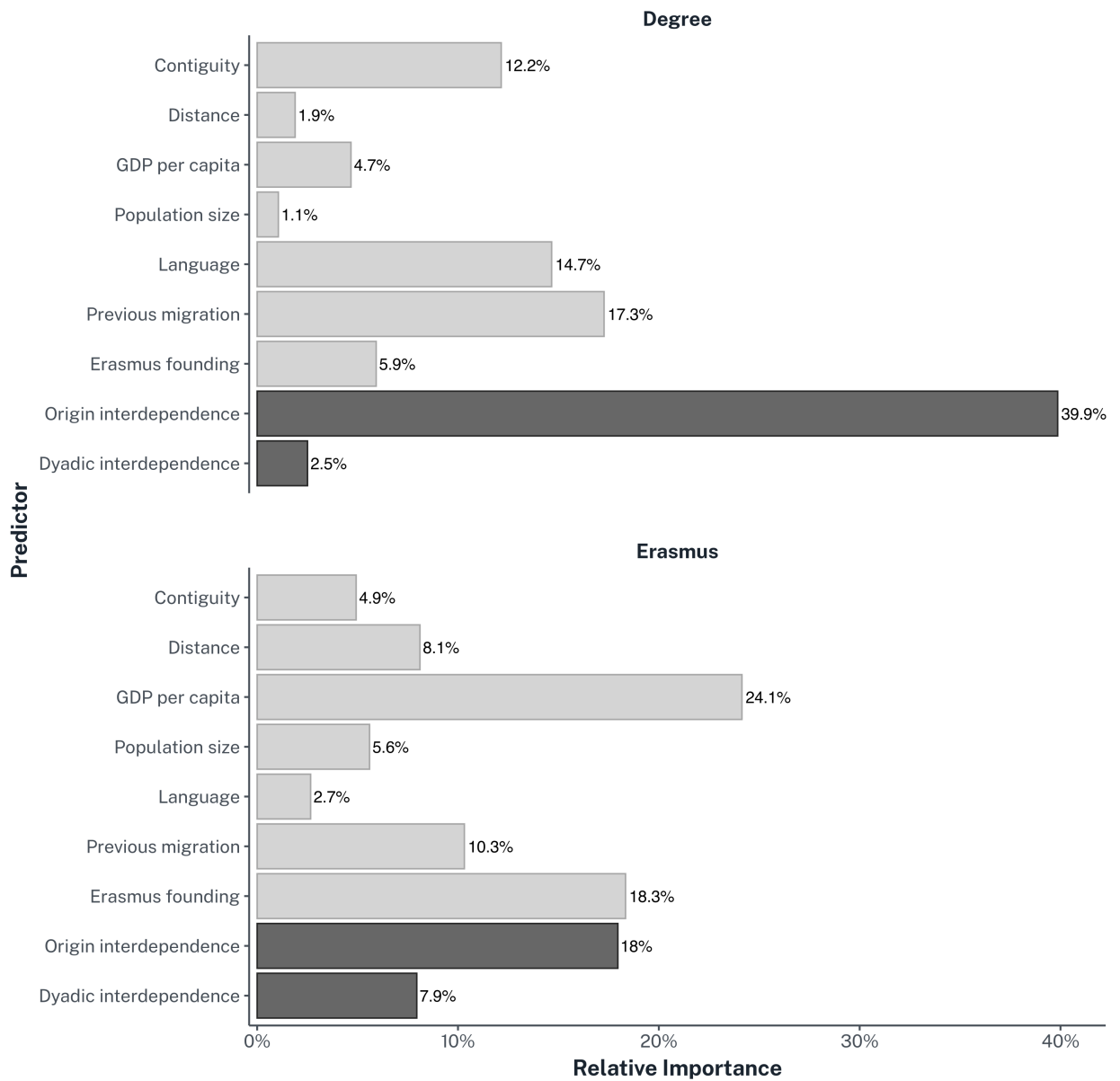


Figure 5.2: Relative explanatory importance (RI) of Model 4 predictors for degree and Erasmus mobility. Relative importance was computed as the average absolute change in individuals' predicted destination-choice probabilities when setting each parameter to zero in turn while holding all other estimates fixed, and then normalised to sum to 100% within each student mobility type.

5.7.2 Measuring Model Performance Across Corridors

Beyond aggregate measures of fit, I assess how well the full model performs at the level of migration corridors, providing an additional diagnostic of whether modelling interdependence improves substantive inference. The heatmaps in Figure 5.3 visualise the corridor-level prediction bias of Model 4: blue cells indicate corridors where the model under-predicts migration (ie the predicted share is below the observed share), while red cells reflect over-predicted corridors. I replicate this figure for the other Models (1–3) in Appendix C.3, Figures C.5 to C.7. As Figure 5.3 shows, prediction bias varies considerably across the two student mobility networks. For degree mobility, the model mispredicts flows involving more peripheral countries, such as flows towards the Czech Republic and Lithuania, but performs better in corridors involving central actors, such as outflows from Italy and inflows to the UK. For Erasmus mobility, the identification of countries as more or less predictable is not as immediate: prediction bias across corridors is distributed less systematically.

For this reason, I investigate three exemplary corridors, highlighted in bold in Figure 5.3: Austria to Germany, Ireland to the Netherlands, and Italy to the UK. These corridors were chosen because they involve countries that have been found to play central roles in the ISM system (Breznik & Skrbinjek, 2020; Hou & Du, 2022; Kondakci et al., 2018; Restaino et al., 2020; Shields, 2013) and because they exhibit prediction bias that provides additional insight into possible shortcomings of the proposed model. Model 4 predicts too little flow from Austria to Germany in the case of degree mobility, but too much in the case of Erasmus mobility. In the corridor from Italy to the UK, Model 4 performs well with respect to degree mobility flows, but overpredicts the number of Erasmus-mobile students, suggesting that Italian students might go to the UK to pursue degrees but not for short-term student exchanges. An intermediate pattern occurs from Ireland to the Netherlands, where Model 4 predicts slightly too much student mobility in both degree and Erasmus.

Corridor Bias Heatmap — Model 4

Bias score compares predicted vs observed corridor shares.

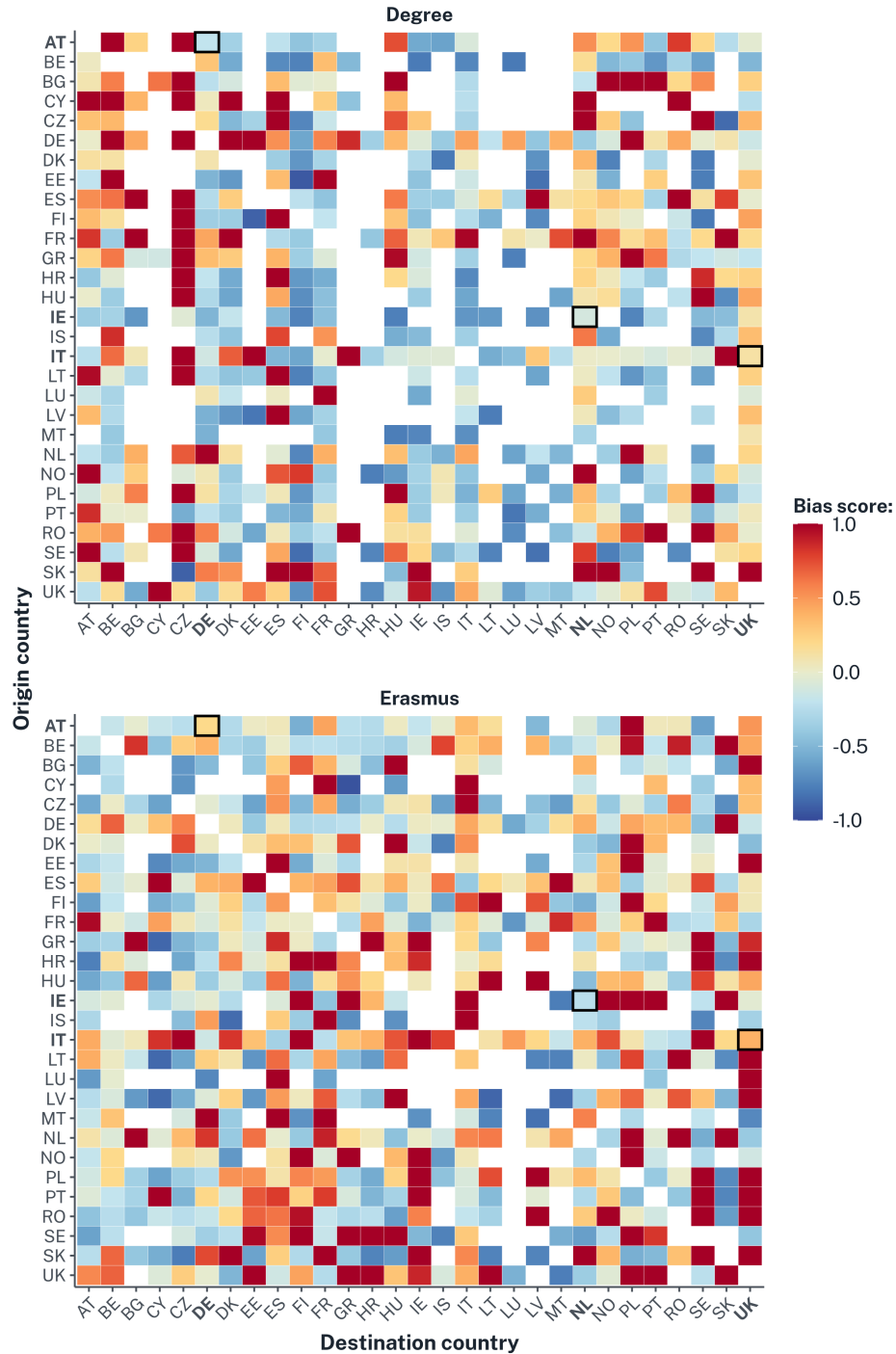


Figure 5.3: Prediction bias of Model 4 for all migration corridors in the sample. Each cell displays the relative deviation between predicted and observed migration flows for a given origin–destination corridor. Blue indicates under-prediction while red indicates over-prediction; the shading reflects the magnitude of the bias metric. Three exemplary corridors (Austria to Germany, Ireland to the Netherlands, and Italy to the UK) are highlighted through bold lettering and borders.

I compare the Model 4 mispredictions for these corridors with those of Model 2, which ignores interdependence effects. I repeatedly simulate the expected flows under these two models, as visualised in Figure 5.4 by mobility type and corridor. The observed numbers of students in each corridor are drawn as dashed lines centred at zero, while the distributions of simulated values are shown as half-violin plots and half-box-plots. Deviations from the observed value are reported in percentages. An extension of this figure to include all Models (1–4) is reported in Appendix C.4, Figure C.8.

In Figure 5.4, it is possible to observe that Model 2 (in light grey) performs systematically worse than Model 4 (in dark grey) for both degree and Erasmus mobility: models assuming independence misestimate flow variability in these corridors, producing biased and overly confident (narrow) simulations. In the degree corridor from Ireland to the Netherlands, for instance, the Model 2 simulations cluster far below the observed flow, with their mean value being over 50% lower than the observed flow.

By contrast, Model 4 produces simulated distributions that are consistently wider and a better approximation of the observed flows, despite the corridor-level prediction bias that emerged in Figure 5.3. This means that modelling origin and dyadic interdependence improves predictive accuracy not only in terms of mean estimates, but also in terms of corridor-specific variability, as seen from the shape and means of the simulation distributions. This improvement is especially pronounced for student flows from Ireland to the Netherlands and for Erasmus exchanges from Austria to Germany, where the Model 4 simulations contain the observed value within the range of possible outcomes. Nonetheless, for certain corridors and mobility types, such as degree mobility from Austria to Germany and Erasmus mobility from Italy to the UK, neither model could generate simulations that include the observed value.

Taken together, the results indicate that conventional predictors of ISM remain relevant and align with theoretical expectations from prior work. However, they do not fully account for the observed spatial organisation of movement. Modelling ISM as

Observed vs Simulated Flows in Selected Corridors

Observed flows centred at zero; half-violins and boxplots show deviations of simulations.

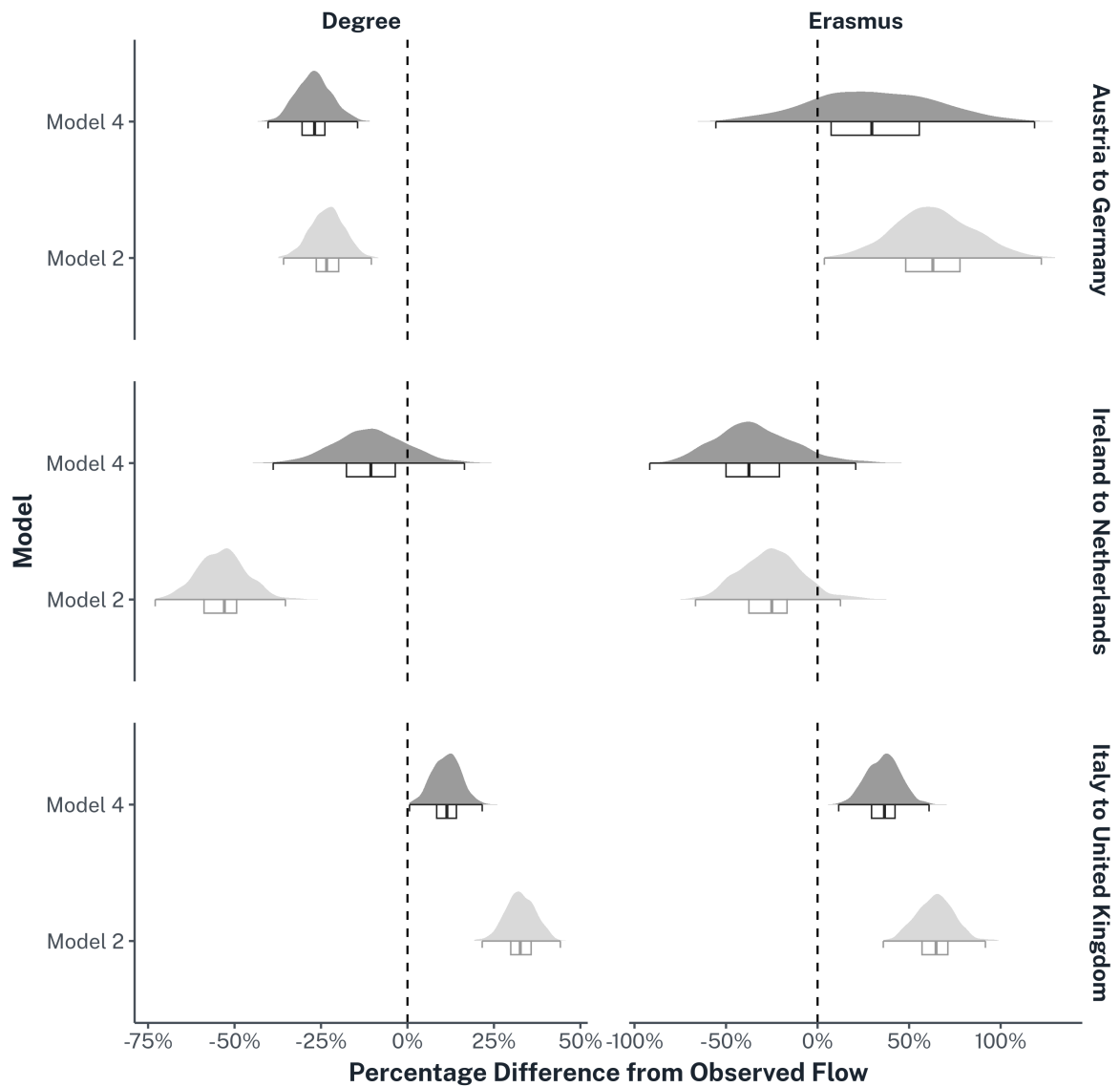


Figure 5.4: Comparison between explanatory models without interdependence (Model 2) and with interdependence (Model 4). For three exemplary corridors (Austria to Germany, Ireland to the Netherlands, Italy to the UK) and by ISM type, 1,000 counterfactual networks were simulated per model; simulated corridor flows are shown relative to observed flows.

networked systems makes it possible to capture the inherently interdependent nature of movement, revealing that, in this context, migration decisions are deeply influenced by concentration (in degree mobility) and systematic asymmetries (in Erasmus mobility).

5.8 Discussion

This chapter set out to bring a core sociological insight—that migration is embedded in relational structures—into direct contact with the dominant empirical practice of modelling flows as conditionally independent once observable drivers are controlled for. I develop an analytical strategy that treats interdependence not as a nuisance to be corrected, but as a structural feature of the migration process that is worthy of estimation and interpretation. Using ELMos, I show how two analytically distinct forms of interdependence, origin and dyadic, can be estimated from aggregate flow data, even when the underlying relational structures are not directly observed. The application to European student mobility demonstrates that interdependence is empirically consequential and a significant organising force shaping where students go.

A central theoretical implication of the results is that the meso-level drivers of migration, that is, the relational structures connecting people to places and to one another, can be operationalised in analyses of aggregate flows through detectable interdependence patterns. Migration theory has long argued that interpersonal networks, institutional infrastructures, and organisations shape destination selection and explain why corridors persist. The challenge has been empirical: those structures are rarely observed comprehensively, and standard models are not designed to represent the interdependence they generate. My proposed approach bridges this gap by estimating origin and dyadic interdependence as endogenous parameters of a statistical model of destination choice, treating them as the empirical signatures of relational structures.

This move is not intended to replace explanations based on individual or contextual attributes. Rather, it complements them by distinguishing two questions that standard models conflate: which contextual attributes make destinations attractive on average, and how migration is organised across corridors as decisions become linked through networks and institutions. In the data used here, country-level predictors explain

meaningful variation, but interdependence accounts for substantial residual structure that is not reducible to bilateral predictors and fixed effects.

Across the two student mobility types, I find that origin interdependence is positive: students from the same origin disproportionately concentrate in the same destinations beyond what shared exposure to the same macro-level conditions can explain. This tendency is consistent with well-established accounts of peer and social influence in ISM, where students from the same origin follow the same migration routes (Beech, 2015; Collins, 2008; Jayadeva, 2024, 2020; Perez-Encinas et al., 2020; Van Mol & Timmerman, 2014). The comparative design of this study clarifies this point. Origin interdependence is substantially stronger in degree mobility, where it achieves a relative explanatory importance of 40% (*vs* 18% in Erasmus), exceeding any conventional predictor. The divergence in the RI of origin interdependence aligns with the institutional contrast between the two ISM types. Where choice sets are relatively unconstrained, as in degree mobility, processes of peer and social influence have greater scope to organise destination selection, generating strong concentration on a limited subset of destinations. Where choice sets are formally bounded by partner lists, as in Erasmus, origin concentration persists but is weaker. This suggests that socially embedded guidance operates within institutional constraints rather than substituting for them.

Conversely, dyadic interdependence in degree mobility appears weakly positive and trivial in relative explanatory importance. Without an institutional infrastructure, there is no strong mechanism linking flows moving in opposite directions of a migration corridor. Rather than a genuine reciprocal exchange, I therefore interpret the weak positive effect in degree mobility as reflecting unmodelled similarities or linkages between origins and destinations, such as active recruitment campaigns and bilateral tuition fee agreements (Jokila et al., 2019; Sin et al., 2021).

The most surprising result concerns the negative dyadic interdependence parameter in Erasmus mobility. Rather than exhibiting the reciprocity that the programme's

bilateral architecture is designed to promote, Erasmus flows display a systematic asymmetry in their direction, net of overall sending and receiving flow volumes. This finding is difficult to reconcile with a simple flow-balancing account of student exchanges. The implication is not that reciprocity norms are absent from the programme, but that they do not translate straightforwardly into symmetric and balanced flows. One plausible interpretation is that HEIs participate in Erasmus partnerships for reasons beyond the reciprocal exchange of students (cf Fumasoli & Rossi, 2021; Gadár et al., 2022). Simply put, sending and receiving HEIs might benefit from one-sided exchange relations in different ways. Sending HEIs may benefit from offering their students attractive destinations and alleviating capacity constraints, while receiving HEIs may benefit from increased international visibility and stronger involvement within European higher education networks. Under this interpretation, what is being exchanged in the Erasmus programme is not simply students but access to networks, symbolic capital, and future institutional leverage (cf Ferreira-Pereira & Mourato Pinto, 2021; Maringe & de Wit, 2016). Negative dyadic interdependence may thus be a system-level trace of strategic positioning within transnational higher education systems. This imbalance resonates with documented core–periphery structures in Erasmus mobility (Breznik & Skrbinjek, 2020; Restaino et al., 2020) and suggests that institutional capacity, prestige, and positional advantages shape the programme in ways that bilateral predictors and formal principles alone cannot capture.

The negative coefficient for Erasmus founding membership complements this reading. I find that exchanges are more likely between founding and non-founding country pairs than within either group, which would be puzzling if one assumes that longer-standing participation should mechanically reinforce exchange ties within the programme's historical core, in a form of path dependence and network segmentation. The findings align more with a strategic view of institutional behaviour: partnerships may be used not only to reinforce established relationships but also to diversify collaborations, build visibility in less saturated corridors, or connect newer members

of the programme to central hubs. Founding-country institutions may function as attractive hosts and network anchors while exhibiting more selective sending patterns, producing the emerging mixed-dyad structure.

Importantly, I remain cautious about the level of analysis when interpreting the Erasmus mobility results. Erasmus agreements are negotiated between individual HEIs, whereas the empirical macro-unit employed in this chapter is the country. The asymmetry documented here should therefore be understood as a system-level signature emerging from heterogeneous institutional agreements within country pairs. The direction of the pattern is nonetheless informative: while many unbalanced institution-level exchanges could, in principle, cancel out and yield balanced origin–destination flows, the dyadic interdependence manifesting at the country level is negative. This indicates that institutional-level asymmetries are themselves systematic rather than randomly distributed, warranting further inquiry into the organisational and symbolic exchanges in which Erasmus mobility is embedded.

Beyond the interpretation of the interdependence parameters, the results of this chapter show that introducing origin and dyadic interdependence changes the estimated effects of conventional bilateral predictors. Effects that appear substantive under independence-based specifications—most strikingly, the positive distance coefficient—attenuate, lose significance, or reverse sign once interdependence is modelled. The implication is substantive: some of what looks like a ‘distance effect’ under conventional models is better understood as the outcome of interdependent destination selection that concentrates flows into specific long-distance corridors. More broadly, where relational structures correlate with macro-level conditions, independence-based models risk conflating the two dimensions.

This finding carries implications beyond the ISM case. If the distance coefficient in the models absorbs concentration behaviour rather than reflecting a genuine preference for long-distance moves, analogous confounding may affect estimates in other migration settings. The broader methodological lesson is that driver-based models and relational

structures are not competing explanations but intertwined: untangling them requires the explicit modelling of interdependence rather than the assumption that controlling for observable bilateral characteristics is sufficient.

Corridor-level diagnostics reinforce this point. The proposed interdependence-aware models improve not only global fit but also the distribution of prediction error across the network space. The corridor-specific misfits are themselves informative: they indicate where distinctive institutional, linguistic, or symbolic factors may be operating and not be captured by standard drivers, or by the two interdependence parameters estimated here, as in the case of degree-mobile Austrians systematically favouring Germany, or Italian students pursuing full degrees but not Erasmus exchanges in the UK. This illustrates a broader advantage of the proposed approach: it generates diagnostics that can guide targeted theorisation and future data collection.

At the same time, several limitations qualify the claims that can be drawn from this study. First, interpreting interdependence parameters as traces of meso-level relational structures presumes that the model accounts reasonably well for macro-level drivers. If important bilateral predictors are omitted, the interdependence parameters may absorb both relational structures and unmodelled contextual heterogeneity. The comparative design in the ISM application mitigates this concern: omitted bilateral characteristics should affect both student mobility types similarly, yet the interdependence parameters diverge in theoretically expected ways. The stability of the interdependence estimates between Model 3 and Model 4 provides further reassurance. The most notable sensitivity arises for the dyadic interdependence parameter in degree mobility, where unmodelled corridor-level similarity could resemble reciprocity, and hence should be interpreted more cautiously. More generally, these considerations point to the importance of specifying a strong baseline model when using interdependence parameters for substantive interpretation. At the same time, they do not undermine the central findings: even where macro-level specifications are imperfect, modelling interdependence would still reduce the bias of conventional predictors.

Second, it should be reinforced that ISM is not inherently organised at the country level. Students choose institutions, programmes, and cities, and Erasmus agreements are negotiated between individual HEIs. Country-level aggregation compresses meaningful within-country variation and may obscure idiosyncratic, institutional-level processes. This limitation is not unique to ISM; analogous issues arise in other forms of international high-skilled migration where individuals move towards particular firms, cities, or sectors rather than countries (Ewers & Dicce, 2018). As noted above, however, the direction of the aggregated patterns is informative. Many asymmetric institution-level flows can produce reciprocal country-level flows, but reciprocal institution-level flows cannot produce the observed country-level asymmetry, given that the countries' general tendency to send and receive migrants is accounted for. The negative dyadic interdependence for Erasmus mobility thus places a lower bound on the true institutional-level asymmetries. Where institutional data are available, future work could apply the ELMo framework to test whether the country-level patterns are reproduced, attenuated, or amplified at finer scales.

Third, the analysis is cross-sectional. The choice of a single year is motivated by the unique characteristics of the empirical case: each year involves a different student cohort, making individual-level longitudinal modelling infeasible. Similarly, lagged flows would introduce autoregressive dependencies that obscure the interpretation of the interdependence parameters. Still, the cross-sectional design means that it is not possible to speak to how interdependence evolves over time: whether origin concentration strengthens as corridors mature, or whether Erasmus asymmetry deepens or corrects as the programme expands. Future research could build on the patterns of synchronic interdependence documented here, where contemporaneous movers influence one another's destination choices, and interrogate the diachronic mechanisms that migration theory has long identified as central to the self-reinforcement of flows. By examining repeated cross-sections, it would be possible to investigate whether the

form and magnitude of interdependence shift or accumulate over time, or whether they respond to contextual changes.

To conclude, the empirical results show that assuming independence in migration decisions is not innocuous: modelling origin and dyadic interdependence reveals organising principles that conventional bilateral predictors do not capture; when omitted, interdependence shifts inference about those predictors and substantively changes our understanding of what drives student moves. I applied the ELMo framework to ISM because this is a setting where relational structures are comparatively well understood, and where a controlled comparison could evaluate whether the model recovers theoretically expected patterns. It does, and this success strengthens confidence in extending the approach to migration contexts where the underlying relational structures are less visible or less theoretically formalised: labour migration organised through recruitment intermediaries, refugee resettlement shaped by diaspora networks, internal migration channelled through housing markets or firm relocations.

More broadly, this chapter speaks to a methodological challenge in sociology that extends beyond migration: how to recover evidence of relational embeddedness from aggregate outcomes when the underlying structures are not directly observed. The statistical framework I employ, and the comparative design I use to validate it, may translate to organisational analysis, stratification research, and other domains where meso-level processes are theorised but rarely measured at scale. If the patterns documented here generalise, interdependence parameters may serve as empirical entry points into relational processes that would otherwise remain obscured by the independence assumptions built into standard models, thereby reshaping what researchers think not only about why people move, but also about how relational structures organise social outcomes more generally.

6 | Conclusion

This thesis began with a simple observation: international migration is not randomly distributed across the globe. Flows concentrate in specific corridors, persist along them, and exhibit regularities in their distribution, concentration, and directionality that migration research has documented but rarely sought to explain. This thesis argued that these regularities constitute migration *structures*, and that explaining human mobility requires tools capable of characterising, evaluating, and modelling these structures. The three empirical chapters sought to develop such tools and demonstrate their value across different migration contexts, geographical scales, and analytical goals.

6.1 Summary of the Empirical Findings

The three empirical chapters engaged with migration structure through distinct but cumulative epistemic modes: description, evaluation, and inference.

The first empirical chapter, *Diversity and Structural Change in International Migration, 1960–2020* (Chapter 3), established that migration structure is empirically tractable and meaningful. By introducing Hill diversity profiles, the chapter showed that the apparent contradiction in the globalisation literature—where some studies find increasing diversity, and others increasing concentration—dissolves once the full distributional shape of the global migration system is examined. Connectivity and evenness have moved in opposite directions since 1960: corridor expansion dominated through the 1980s, while flow redistribution toward smaller corridors dominated from the 1990s onward. The within- and between-country decomposition further revealed

that rising global evenness coexists with a within-country diversification of migration portfolios and a progressive convergence towards the same dominant partners. These findings demonstrated that characterising migration structure requires attention to the full distribution of flows rather than to any single summary index, and established the descriptive foundation for asking whether theories of migration can account for the structures they purport to explain.

The second empirical chapter, *A Simulation Approach for the Spatial Testing of Migration Theories* (Chapter 4), took up this question directly. It developed a simulation-based framework for evaluating migration theories not only by their explanatory power but also by their spatial accuracy: whether the migration systems they imply resemble those observed in the real world. Applying this framework to intra-European migration between 2002 and 2021, the chapter found that both the gravity model and Migration Systems Theory (MST) achieve high explanatory power under conventional goodness-of-fit metrics, yet fail to reproduce key spatial features of the European migration system: most notably, the high reciprocity of flows and the distinctive migration profiles of Eastern and Northern European countries. This result indicates that accounting for migration volume is not the same as accounting for structure.

The third empirical chapter, *The Interdependence of Migration: How Relational Structures Organise Movement* (Chapter 5), responded to the theoretical shortcomings identified in Chapter 4 by challenging a foundational assumption in quantitative migration models: the conditional independence of migration decisions. Using a state-of-the-art statistical network model applied to European student mobility, the chapter estimated two forms of interdependence: origin interdependence, which captures the tendency for migrants from the same origin to concentrate in the same destinations, and dyadic interdependence, which captures systematic associations between flows occurring in opposite directions within a migration corridor. The comparative design contrasted degree mobility with Erasmus exchanges and provided leverage for interpreting these parameters. Origin interdependence was positive and

substantial in both mobility types, but considerably stronger in degree mobility, where it accounted for 40% of relative explanatory importance. Dyadic interdependence diverged: weakly positive in degree mobility, but significantly negative in Erasmus mobility, pointing to systematic asymmetries despite the programme's stated principle of balanced exchange. Accounting for interdependence also reshaped conclusions about conventional predictors: the effect of geographic distance, for instance, attenuated or reversed once interdependence was modelled explicitly.

6.2 Contributions

Viewed as an integrated whole, this thesis makes contributions in three main areas: conceptual, methodological, and substantive.

A first *conceptual* contribution is to establish migration structure as an analytically distinct object from migration volume and to use it as an integrating concept for two research traditions that have developed largely in parallel. Driver-oriented research has sought to explain why some corridors carry more migrants than others; pattern-oriented research has described how flows are distributed across space. The divide between these traditions is not merely disciplinary: it reflects a substantive gap in how migration theories are developed and evaluated, as spatial accuracy is seldom considered when assessing theoretical and model validity. By treating structure as the common object, this thesis makes it possible to ask not only whether drivers explain volume, but also whether they explain the spatial configurations that migration systems exhibit.

A further conceptual contribution of the thesis is the understanding of interdependence as an empirically tractable property of aggregate flows: a statistical regularity indicating that migration decisions are interlinked beyond what bilateral characteristics can explain. This understanding also clarifies the relationship between interdependence and migration mechanisms. Migration mechanisms—peer influence,

chain migration, cumulative causation, organisational channelling—are the *relational processes* that generate this interdependence. The distinction matters because interdependence can be estimated from aggregate data even when the mechanisms producing it are not directly observed, and its parameters carry a dual interpretation: at the micro level, they describe behavioural tendencies in destination choice; at the system level, they correspond to network statistics capturing concentration and reciprocity across the migration system.

The thesis's intended *methodological* contribution is to demonstrate that different analytical tasks—description, evaluation, inference—require different tools, and that these tools can be deployed in a cumulative sequence. Hill diversity profiles provide a principled framework for characterising the distributional properties of migration systems, making explicit the weighting choices that single indices conceal. The simulation-based spatial tests generate counterfactual migration systems from fitted models and compare their spatial features to those of the observed system, enabling direct assessment of whether theories that explain volume also reproduce structure. The statistical network model estimates interdependence parameters alongside conventional predictors, quantifying what standard models may miss and diagnosing how their estimates may be biased. Each framework is portable to other migration contexts and, in some cases, to domains of social research where relational structures are theorised but rarely measured at scale. More broadly, the empirical chapters employ methods characteristic of computational social science, in particular simulations and network approaches, and demonstrate that these methods are not merely technical refinements but necessary tools that reshape our understanding of migration processes.

The *substantive* contribution lies in three findings that would not be accessible through approaches focused on migration volume alone, and in what they imply when read together. First, the distributional framework reveals that the globalisation of migration proceeded through two independent structural processes operating at different historical moments (corridor expansion and flow redistribution) and that the

apparent paradox in the literature dissolves once migration diversity is understood as a distributional property rather than a single summary measure. The Hill framework has value beyond this specific application. International migration is often discussed in aggregate terms—total inflows, net balances, share of foreign-born population—but such figures obscure the spatial organisation that shapes where migration’s effects concentrate. Two countries may receive comparable numbers of immigrants yet differ markedly in whether those arrivals come from many origins or few, and in whether the same corridors connect them to the rest of the world. Similarly, research on immigrant integration, social cohesion, and the societal consequences of diversity routinely relies on single-point indices of immigration or emigration diversity, yet different indices weight dominant and peripheral corridors differently, and can rank the same country differently depending on which part of the distribution they emphasise. Whether a society is ‘diverse’ because it receives small numbers of people from many origins, or because it hosts equally sized communities from a broad array of countries, is a question that single indices cannot distinguish, but Hill profiles can. Making this distinction available to integration research and to the policy frameworks that depend on it is a direct contribution of the distributional approach developed here.

Second, the simulation-based framework demonstrates that established migration theories fail to reproduce key spatial features of the European migration system despite achieving high explanatory power. The implication is that it is not sufficient to identify which drivers matter—economic differentials, geographic distance, colonial ties—if models incorporating those drivers cannot reconstruct the observed spatial configurations. This finding is particularly relevant in settings where policy research and policymaking rely on models to forecast and anticipate migration patterns.

Third, one possible response to the limits of driver-based models would be to pursue ever-richer specifications, incorporating additional predictors in the expectation that the residual spatial misfit will eventually disappear. The findings of Chapter 5, however, suggest that this strategy may be insufficient on its own. Origin and

dyadic interdependence account for substantial shares of explanatory importance, and omitting them distorts inference about conventional predictors: what appears to be a distance effect under independence-based models is partly an artefact of concentrated destination selection that standard specifications cannot represent. The direct consequence is that spatial accuracy requires not only better-specified drivers but also the explicit modelling of interdependence as a primary organising force of the spatial distribution of migration. For policymakers seeking to understand migration systems rather than simply count migrants, attending to structure and interdependence would therefore offer a more complete picture of how mobile the world truly is and how that mobility unfolds.

6.3 Limitations

Several limitations qualify the claims of this thesis. Some are specific to individual chapters; others concern the thesis as a whole.

At the chapter level, Chapter 3 is descriptive throughout: it establishes what happened to migration diversity and when, but does not identify the mechanisms that produced observed changes. Its ten-year aggregation limits temporal resolution, and the reliance on flow estimates rather than registered counts introduces uncertainty that cannot be fully eliminated, even if partially addressed through robustness checks. Chapter 4 simulates dyadic flows independently, abstracting from individual-level destination choice: while the approach can be extended to micro-level simulations, including Agent-Based Models (ABMs) and microsimulations, the present application does not adjudicate whether reciprocity misfits reflect individual behaviour or aggregate-level omitted variables. The chapter's focus on intra-European migration also excludes corridors from the Global South to the Global North, where the hierarchical expectations of MST might perform differently. Finally, Chapter 5 operates at the country level, compressing meaningful within-country variation and potentially

obscuring institutional-level processes. Its cross-sectional design, although pertinent to the migration case at hand, also precludes the analysis of how interdependence evolves over time. The interdependence parameters in the student mobility application capture the behaviour of other students in the network within a single year; they do not estimate the accumulated effects of prior migration, such as those arising from cumulative causation or the gradual consolidation of diaspora networks. This means that the chapter speaks to synchronic interdependence—how migration decisions are linked within the same period—but not to the diachronic mechanisms that migration theory has long identified as instrumental to the continuation and self-reinforcement of movement.

At the thesis level, the most significant structural limitation concerns scope and integration across the empirical chapters. Because the chapters were designed to stand alone as self-contained papers, each considers a distinct migration case: global bilateral flows over six decades, intra-European migration over two decades, and intra-European student mobility in a single year. This progression was deliberate: as inferential ambition increased, so did the data precision required to support it. The incremental design, however, means that the chapters do not share a common empirical object, which limits the degree to which findings can be formally integrated. Chapter 4, for instance, finds that observed intra-European migration is more reciprocal than gravity or MST would predict; Chapter 5 introduces interdependence parameters that could, in principle, account for this misfit. Whether modelling interdependence would improve spatial accuracy in Chapter 4's simulations remains an open question, and one that I hope to address in the near future.

A second thesis-level limitation concerns the indirect detection of migration mechanisms. Migration mechanisms—peer influence, chain migration, cumulative causation, organisational channelling—are central to the thesis's theoretical architecture: it is precisely the expectation that such mechanisms produce emergent structures that motivates the need to go beyond explanations of volume alone. In this sense, mechan-

isms permeate the thesis, but no empirical chapter observes them directly. Chapter 3 documents patterns whose timing and form are consistent with structural transformations, but cannot identify the specific mechanisms behind these transformations. Chapter 4 demonstrates that driver-based models fail to reproduce observed structure, a finding whose theoretical significance depends on the premise that unmodelled mechanisms are responsible for the misfit. Chapter 5 comes closest to empirical engagement: its interdependence parameters are estimated within a micro-founded statistical framework and are consistent with theorised relational processes. Even here, however, the model infers the presence of mechanisms from their aggregate traces, without directly measuring the mechanisms themselves.

The challenge of measuring mechanisms is not unique to this thesis; it is a pervasive challenge in sociology, and one that analytical sociology has long grappled with (Hedström, 2005; Hedström & Bearman, 2009; Hedström & Ylikoski, 2010; Ylikoski, 2011). The aim of the present work is to show that mechanisms can be brought into empirical contact with aggregate data through their structural signatures, even when they cannot be observed at the level at which they operate. This is a weaker claim than direct mechanistic identification, but not an empty one: it provides ‘counterfactual dependence’ that constitutes explanatory relevance even if the underlying causal processes remain incompletely specified (Schmid, 2011; Ylikoski, 2011). Nonetheless, connecting the macro-level traces documented here to direct micro- and meso-level evidence of the mechanisms that generate them would strengthen the thesis’s central claims and thus represents a natural extension of this work.

Relatedly, all three empirical chapters operate at the macro level of aggregate bilateral flows, leaving the micro level largely unattended. Neither the simulations in Chapter 4 nor the statistical network model in Chapter 5 operates at the individual level or incorporate migrants’ characteristics, despite being in principle implementable at the micro level. This strategy reflects the thesis’s deliberate focus on establishing the system-level importance of migration structure, but it leaves important blind spots. The

absence of individual-level variation means that the thesis cannot address whether the patterns it documents emerge uniformly from migrant populations or are concentrated among specific subgroups. Migration theory suggests that migration mechanisms are highly heterogeneous, operating with different intensity for different populations, at different stages of the migration process, and under different institutional conditions (see Beine & Salomone, 2013; Curran & Saguy, 2001; Garip, 2012, 2019; Garip & Asad, 2016; Heering et al., 2004; Liu, 2013; Toma & Vause, 2014; Zhao & Qu, 2022). The comparative design of Chapter 5 demonstrated that interdependence parameters can distinguish between institutionally different contexts, but further effect decompositions by migrant characteristics, such as age and gender, would better align the system-level estimates with mechanism-specific heterogeneity.

A final limitation concerns generalisation. The increasing data demands across chapters—from estimated global flows to harmonised regional flows to register-based administrative counts—mean that the thesis’s most inferentially ambitious methods require data infrastructures that remain unevenly developed across regions and migration types. The statistical network model was applied to European student mobility because this setting provides reliable bilateral data, consistent migrant definitions, well-understood institutional structures, and sufficient corridor variation to identify interdependence parameters. These conditions were necessary to evaluate both the reliability and the substantive value of the method; in this sense, the stringent empirical case serves as a proof of concept. Nonetheless, the specific characteristics of student mobility, which set it apart from other migration cases, might limit the portability of the findings to different contexts.

6.4 Areas for Future Research

The limitations identified above point to several productive directions for future work, each following from a specific gap that the thesis exposes but does not close.

The most immediate extension concerns the integration of the methods developed across the empirical chapters. The spatial tests introduced in Chapter 4 could be applied to the global flow data of Chapter 3 to investigate which migration theories, if any, best account for the changing patterns of migration diversity. More directly, the interdependence parameters estimated in Chapter 5 could be incorporated into the simulation-based framework of Chapter 4. Dyadic interdependence operationalises a form of reciprocity at the system level; if the reciprocity misfits in Chapter 4 stem partly from the independence assumption built into the gravity and MST specifications, explicitly modelling interdependence should reduce them. Pursuing this integration would test whether the theoretical diagnosis offered by Chapter 5 translates into measurable improvements in spatial accuracy.

A second direction follows from the absence of micro-level analysis. In Chapter 4, replacing the macro-level simulation strategy with ABMs or microsimulations would allow the spatial tests to evaluate theories that operate at the individual level, such as network theory or the new economics of labour migration. In Chapter 5, the ELMo framework could be extended to incorporate individual-level predictors, such as age and gender, alongside the interdependence parameters. Interacting individual attributes with the endogenous parameters would open a further set of research questions: whether women and men differ in the extent to which their destination choices are shaped by the behaviour of their peers, or whether social influence varies across stages of the life course. Such interactions would also begin to decompose aggregate interdependence signals into subgroup-specific components, connecting the system-level parameters to the mechanism-specific heterogeneity. If, for instance, origin concentration is substantially stronger for female migrants in a context where women rely more heavily on established networks for information and support, this would be consistent with direct social facilitation as the dominant mechanism for that subgroup; if concentration is equally strong for men and women, network externalities operating through community-level infrastructures become a more plausible account. The thesis

establishes that interdependence matters at the system level; disaggregating it by migrant characteristics is the natural next step toward identifying which mechanisms may drive it.

A third direction concerns the mechanisms that the thesis theorises but does not observe. The decomposition strategy outlined above—interacting interdependence parameters with individual attributes—represents one route toward narrowing the set of plausible mechanisms. A complementary route would combine the macro-level frameworks developed here with individual-level data on social ties, institutional affiliations, and decision-making processes. Datasets that contain both detailed micro-level information and sufficient bilateral coverage to estimate interdependence parameters would allow researchers to test whether the system-level traces recovered from aggregate flows correspond to the specific relational processes that migration theory identifies. Such an approach would move the analysis toward a more sophisticated multi-level explanatory architecture that formally specifies how individual decisions, shaped by relational structures, produce the emergent system-level configurations that this thesis documents.

Finally, future research could extend the proposed frameworks to migration settings and populations beyond those examined here. For instance, studies of internal migration may exploit finer geographical resolution and richer administrative data and implement the methods in a new testing ground. The statistical network model may also translate to domains beyond migration, such as social mobility, wherever meso-level relational processes are theorised but rarely measured at scale.

In a broader perspective, attending to migration structure—describing it carefully, evaluating whether theories reproduce it, and modelling the relational processes that shape it—yields insights not accessible through approaches focused on migration volume alone. The gap between explaining volume and reproducing structure is not a technical inconvenience but a substantive limitation in how migration research has framed its explanatory ambitions. The tools introduced here to characterise migration

diversity, test spatial accuracy, and model interdependence are offered as contributions toward a reorientation in that framing. If this thesis helps establish migration structure as a first-class analytical object and provides instruments for studying it, it will have achieved its purpose.

Bibliography

- Abbott, A., & Silles, M. (2016). Determinants of International Student Migration. *The World Economy*, 39(5), 621–635.
- Abel, G. J. (2018). Estimates of Global Bilateral Migration Flows by Gender Between 1960 and 2015. *International Migration Review*, 52(3), 809–852.
- Abel, G. J. (2025). *migest: Methods for the Indirect Estimation of Bilateral Migration* [R package version 2.0.5]. <http://guyabel.github.io/migest/>
- Abel, G. J., & Cohen, J. E. (2019). Bilateral International Migration Flow Estimates for 200 Countries. *Scientific Data*, 6(82).
- Abel, G. J., & Cohen, J. E. (2022). Bilateral International Migration Flow Estimates Updated and Refined by Sex. *Scientific Data*, 9(173).
- Abel, G. J., DeWaard, J., Ha, J. T., & Almquist, Z. W. (2021). The Form and Evolution of International Migration Networks, 1990–2015. *Population, Space and Place*, 27(3).
- Abel, G. J., & Sander, N. (2014). Quantifying Global International Migration Flows. *Science*, 343(6178), 1520–1522.
- Åberg, Y. (2000). Individual Social Action and Macro Level Dynamics: A Formal Theoretical Model. *Acta Sociologica*, 43, 193–205.
- Adhvaryu, A., Gauthier, J.-F., Nyshadham, A., & Tamayo, J. (2024). Absenteeism, Productivity, and Relational Contracts Inside the Firm. *Journal of the European Economic Association*, 22(4), 1628–1677.
- Akbari, H. (2021). Exploratory Social-Spatial Network Analysis of Global Migration Structure. *Social Networks*, 64, 181–193.
- Aksoy, O., & Yildirim, S. (2024). A Model of Dynamic Flows: Explaining Turkey's Interprovincial Migration. *Sociological Methodology*, 54(1), 52–91.
- Allison, P. D. (2014). Measures of Fit for Logistic Regression. *Proceedings of the SAS Global Forum 2014 Conference*.

- Alon, S., & Gelbgiser, D. (2011). The Female Advantage in College Academic Achievements and Horizontal Sex Segregation. *Social Science Research*, 40(1), 107–119.
- Andresen, M., Bergdolt, F., Margenfeld, J., & Dickmann, M. (2014). Addressing International Mobility Confusion – Developing Definitions and Differentiations for Self-Initiated and Assigned Expatriates as Well as Migrants. *The International Journal of Human Resource Management*, 25(16), 2295–2318.
- Anniste, K., & Tammaru, T. (2014). Ethnic Differences in Integration Levels and Return Migration Intentions: A Study of Estonian Migrants in Finland. *Demographic Research*, 30, 377–412.
- Arango, J. (2000). Explaining Migration: A Critical View. *International Social Science Journal*, 52(165), 283–296.
- Azose, J. J., & Raftery, A. E. (2019). Estimation of Emigration, Return Migration, and Transit Migration Between All Pairs of Countries. *Proceedings of the National Academy of Sciences*, 116(1), 116–122.
- Bakewell, O. (2014). Relaunching Migration Systems. *Migration Studies*, 2(3), 300–318.
- Bakewell, O. (2024). Migration Systems. In G. Sciortino, M. Cvajner & P. J. Kivisto (Eds.), *Research Handbook on the Sociology of Migration* (pp. 86–95). Edward Elgar Publishing.
- Bakewell, O. (2010). Some Reflections on Structure and Agency in Migration Theory. *Journal of Ethnic and Migration Studies*, 36(10), 1689–1708.
- Bakewell, O., de Haas, H., & Kubal, A. (2012). Migration Systems, Pioneer Migrants and the Role of Agency. *Journal of Critical Realism*, 11(4), 413–437.
- Bakewell, O., Engbersen, G., Fonseca, M. L., & Horst, C. (Eds.). (2016). *Beyond Networks: Feedback in International Migration*. Palgrave Macmillan UK.
- Bakewell, O., & Jolivet, D. (2016). Broadcasting Migration Outcomes. In O. Bakewell, G. Engbersen, M. L. Fonseca & C. Horst (Eds.), *Beyond Networks: Feedback in International Migration Processes* (pp. 183–204). Palgrave Macmillan UK.
- Bakewell, O., Kubal, A., & Pereira, S. (2016). Introduction: Feedback in Migration Processes. In O. Bakewell, G. Engbersen, M. L. Fonseca & C. Horst (Eds.), *Beyond Networks: Feedback in International Migration Processes* (pp. 1–17). Palgrave Macmillan UK.

- Barnett, G. A., Lee, M., Jiang, K., & Park, H. W. (2016). The Flow of International Students from a Macro Perspective: A Network Analysis. *Compare: A Journal of Comparative and International Education*, 46(4), 533–559.
- Barnett, G. A., & Nam, Y. (2024). A Network Analysis of International Migration: Longitudinal Trends and Antecedent Factors Predicting Migration. *Global Networks*, 24(2), e12455.
- Batista, C., & McKenzie, D. (2023). Testing Classic Theories of Migration in the Lab. *Journal of International Economics*, 145.
- Beech, S. E. (2015). International Student Mobility: The Role of Social Networks. *Social & Cultural Geography*, 16(3), 332–350.
- Beerkens, M., Souto-Otero, M., de Wit, H., & Huisman, J. (2016). Similar Students and Different Countries? An Analysis of the Barriers and Drivers for Erasmus Participation in Seven Countries. *Journal of Studies in International Education*, 20(2), 184–204.
- Beine, M., Bertoli, S., & Fernández-Huertas Moraga, J. (2016). A Practitioners' Guide to Gravity Models of International Migration. *The World Economy*, 39(4), 496–512.
- Beine, M., Bierlaire, M., & Docquier, F. (2025). New York, Abu Dhabi, London, or Stay at Home? Using a Cross-Nested Logit Model to Identify Complex Substitution Patterns in Migration. *Journal of Economic Geography*, 25(5), 685–712.
- Beine, M., Noël, R., & Ragot, L. (2014). Determinants of the International Mobility of Students. *Economics of Education Review*, 41, 40–54.
- Beine, M., & Salomone, S. (2013). Network Effects in International Migration: Education versus Gender. *The Scandinavian Journal of Economics*, 115(2), 354–380.
- Bell, M., Blake, M., Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J., & Hugo, G. (2002). Cross-National Comparison of Internal Migration: Issues and Measures. *Journal of the Royal Statistical Society*, 165, 435–464.
- Bertoli, S., & Fernández-Huertas Moraga, J. (2013). Multilateral Resistance to Migration. *Journal of Development Economics*, 102, 79–100.
- Bijak, J. (2022). *Towards Bayesian Model-Based Demography: Agency, Complexity and Uncertainty in Migration Studies* (Vol. 17). Springer International Publishing.

- Bijak, J., Courgeau, D., Franck, R., & Silverman, E. (2018). Modelling in Demography: From Statistics to Simulations. In *Methodological Investigations in Agent-Based Modelling* (pp. 167–187). Springer International Publishing.
- Bilecen, B., Gamper, M., & Lubbers, M. J. (2018). The Missing Link: Social Network Analysis in Migration and Transnationalism. *Social Networks*, 53, 1–3.
- Block, P. (2023). Understanding the Self-Organization of Occupational Sex Segregation with Mobility Networks. *Social Networks*, 73, 42–50.
- Block, P., Koskinen, J., Hollway, J., Steglich, C., & Stadtfeld, C. (2018). Change We Can Believe In: Comparing Longitudinal Network Models on Consistency, Interpretability and Predictive Power. *Social Networks*, 52, 180–191.
- Block, P., Stadtfeld, C., & Robins, G. (2022). A Statistical Model for the Analysis of Mobility Tables as Weighted Networks with an Application to Faculty Hiring Networks. *Social Networks*, 68, 264–278.
- Bohra, P., & Massey, D. S. (2009). Processes of Internal and International Migration from Chitwan, Nepal. *International Migration Review*, 43(3), 621–651.
- Botezat, A., & Ramos, R. (2020). Physicians’ Brain Drain – A Gravity Model of Migration Flows. *Globalization and Health*, 16(1).
- Boyle, P., Halfacree, K., & Robinson, V. (1998). *Exploring Contemporary Migration* (1st ed.). Routledge.
- Brettell, C. B., & Hollifield, J. F. (2022). Introduction. In C. B. Brettell & J. F. Hollifield (Eds.), *Migration Theory: Talking Across Disciplines* (4th ed., pp. 1–43). Routledge.
- Breznik, K., & Skrbinjek, V. (2020). Erasmus Student Mobility Flows. *European Journal of Education*, 55(1), 105–117.
- Chao, A., Chiu, C.-H., & Jost, L. (2010). Phylogenetic Diversity Measures Based on Hill Numbers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1558), 3599–3609.
- Charles-Edwards, E., Bernard, A., Rowe, F., & Abel, G. J. (2025). International Migration and Development: The Changing Impact of Migration on Redistributing Global Population. *International Migration Review*, 59(2), 850–866.

- Choudaha, R. (2017). Three Waves of International Student Mobility (1999–2020). *Studies in Higher Education*, 42(5), 825–832.
- Collins, F. L. (2008). Bridges to Learning: International Student Mobilities, Education Agencies and Inter-Personal Networks. *Global Networks*, 8(4), 398–417.
- Colombo, A. D., & Dalla-Zuanna, G. (2019). Immigration Italian Style, 1977–2018. *Population and Development Review*, 45(3), 585–615.
- Colombo, A. D., & Sciortino, G. (2004). Italian Immigration: The Origins, Nature and Evolution of Italy's Migratory Systems. *Journal of Modern Italian Studies*, 9(1), 49–70.
- Conte, M., Cotterlaz, P., & Mayer, T. (2021). The CEPII Gravity Database. *CEPII Working Paper*. Retrieved July 6, 2025, from http://www.cepii.fr/DATA_DOWNLOAD/gravity/doc/Gravity_documentation.pdf
- Cranston, S., Schapendonk, J., & Spaan, E. (2018). New directions in exploring the migration industries: Introduction to special issue. *Journal of Ethnic and Migration Studies*, 44(4), 543–557.
- Curran, S. R., & Saguy, A. C. (2001). Migration and Cultural Change: A Role for Gender and Social Networks? *Journal of International Women's Studies*, 2(3), 54–77.
- Czaika, M., Bohnet, H., & Soto-Nishimura, A. (2025). Spatial Dependence of European Immigration Flows. *Population and Development Review*, padr.70011.
- Czaika, M., & de Haas, H. (2014). The Globalization of Migration: Has the World Become More Migratory? *International Migration Review*, 48(2), 283–323.
- Czaika, M., & Reinprecht, C. (2022). Migration Drivers: Why Do People Migrate? In P. Scholten (Ed.), *Introduction to Migration Studies* (pp. 49–82). Springer International Publishing.
- Daly, A. J., Baetens, J. M., & De Baets, B. (2018). Ecological Diversity: Measuring the Unmeasurable. *Mathematics*, 6(7).
- Danchev, V., & Porter, M. A. (2021). Migration Networks: Applications of Network Analysis to Macroscale Migration Patterns. In M. McAuliffe (Ed.), *Research Handbook on International Migration and Digital Technology* (pp. 70–90). Edward Elgar Publishing.
- Danchev, V., & Porter, M. A. (2018). Neither Global nor Local: Heterogeneous Connectivity in Spatial Network Structures of World Migration. *Social Networks*, 53, 4–19.

- Dańko, M. J. (2023). HMigD I App: the Human Migration Database I. Retrieved March 11, 2025, from https://maciej-jan-danko.shinyapps.io/HMigD_Shiny_App_I/
- Dańko, M. J., Wiśniowski, A., Jasilionis, D., Jdanov, D. A., & Zagheni, E. (2024). Assessing the Quality of Data on International Migration Flows in Europe: The Case of Undercounting. *Migration Studies*, 12(2), 1–21.
- Davis, K. F., D’Odorico, P., Laio, F., & Ridolfi, L. (2013). Global Spatio-Temporal Patterns in Human Migration: A Complex Network Perspective. *PLoS ONE*, 8(1), e53723.
- de Wit, H., Ferencz, I., & Rumbley, L. E. (2012). International Student Mobility: European and US Perspectives. *Perspectives: Policy and Practice in Higher Education*, 1–7.
- Debonneville, J. (2021). An Organizational Approach to the Philippine Migration Industry: Recruiting, Matching and Tailoring Migrant Domestic Workers. *Comparative Migration Studies*, 9(12).
- de Haas, H. (2021). A Theory of Migration: The Aspirations-Capabilities Framework. *Comparative Migration Studies*, 9(8).
- de Haas, H. (2010). The Internal Dynamics of Migration Processes: A Theoretical Inquiry. *Journal of Ethnic and Migration Studies*, 36(10), 1587–1617.
- de Haas, H., Castles, S., & Miller, M. J. (2022). *The Age of Migration: International Population Movements in the Modern World* (6th ed.). Bloomsbury Academic.
- de Haas, H., Czaika, M., Flahaux, M.-L., Mahendra, E., Natter, K., Vezzoli, S., & Villares-Varela, M. (2019). International Migration: Trends, Determinants, and Policy Effects. *Population and Development Review*, 45(4), 885–922.
- Denisenko, M., Strozza, S., & Light, M. (Eds.). (2020). *Migration from the Newly Independent States: 25 Years After the Collapse of the USSR*. Springer International Publishing.
- Destatis. (2023). Die Datenbank des Statistischen Bundesamtes, GENESIS-Online, Migration Statistics, 12711. Retrieved September 22, 2023, from www-genesis.destatis.de/genesis/online?sequenz=statistikTabellen&selectionname=12711&language=en#abreadcrumb
- DeWaard, J., & Ha, J. T. (2019). Resituating Relaunched Migration Systems as Emergent Entities Manifested in Geographic Structures. *Migration Studies*, 7(1), 39–58.

- DeWaard, J., Kim, K., & Raymer, J. (2012). Migration Systems in Europe: Evidence from Harmonized Flow Data. *Demography*, 49(4), 1307–1333.
- Dolfin, S., & Genicot, G. (2010). What Do Networks Do? The Role of Networks on Migration and “Coyote” Use. *Review of Development Economics*, 14(2), 343–359.
- Drouhot, L. G., Deutschmann, E., Zuccotti, C. V., & Zagheni, E. (2023). Computational Approaches to Migration and Integration Research: Promises and Challenges. *Journal of Ethnic and Migration Studies*, 49(2), 389–407.
- Edelmann, A., Wolff, T., Montagne, D., & Bail, C. A. (2020). Computational Social Science and Sociology. *Annual Review of Sociology*, 46(1), 61–81.
- Elhorst, J. (2013). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer.
- Engbersen, G., Snel, E., & Esteves, A. (2016). Migration Mechanisms of the Middle Range: On the Concept of Reverse Cumulative Causation. In O. Bakewell, G. Engbersen, M. L. Fonseca & C. Horst (Eds.), *Beyond Networks* (pp. 205–230). Palgrave Macmillan.
- Engbersen, G., Snel, E., & Horst, C. (2016). Beyond Networks: Insights on Feedback and Mechanisms of the Middle Range. In O. Bakewell, G. Engbersen, M. L. Fonseca & C. Horst (Eds.), *Beyond Networks: Feedback in International Migration Processes* (pp. 231–246). Palgrave Macmillan UK.
- eQuATIC Project & European University Foundation. (2019). *The Use of Data for the Internationalisation of Higher Education* (Research Paper). European University Foundation. https://uni-foundation.eu/uploads/2019_eQuATIC_research_paper.pdf
- European Commission. (2017). From Erasmus to Erasmus+: A Story of 30 Years [EC MEMO/17/83]. https://ec.europa.eu/commission/presscorner/detail/en/MEMO_17_83
- European Commission. (2023). Guidance on the Right of Free Movement of EU Citizens and Their Families. *Official Journal of the European Union*, 2023/1392. Retrieved November 6, 2024, from <http://data.europa.eu/eli/C/2023/1392/oj>
- European Commission. (2015). Towards a Comprehensive European Migration Policy: 20 Years of EU Action [Fact Sheet, Brussels]. Retrieved July 6, 2025, from https://ec.europa.eu/commission/presscorner/api/files/document/print/en/memo_15_4544/MEMO_15_4544_EN.pdf

- European Commission, Directorate-General for Education, Youth, Sport and Culture. (2021). Erasmus Mobility Statistics 2014–2019 [Data set]. Retrieved July 6, 2023, from <http://data.europa.eu/88u/dataset/erasmus-mobility-statistics-2014-2019-v2>
- Eurostat. (2023a). Emigration by Age Group, Sex and Country of Next Usual Residence. Retrieved September 22, 2023, from https://ec.europa.eu/eurostat/en/web/products-datasets/-/MIGR_EMI3NXT
- Eurostat. (2023b). Immigration by Age Group, Sex and Country of Previous Residence. Retrieved September 22, 2023, from https://ec.europa.eu/eurostat/en/web/products-datasets/-/MIGR_IMM5PRV
- Ewers, M. C., & Dicce, R. (2018, February 15). High-Skilled Migration and the Attractiveness of Cities. In M. Czaika (Ed.), *High-Skilled Migration: Drivers and Policies* (Vol. 1). Oxford University Press.
- Faist, T. (1997). The Crucial Meso-Level. In T. Hammar (Ed.), *International Migration, Immobility and Development: Multidisciplinary Perspectives*. Taylor & Francis.
- Faist, T. (2000). *The Volume and Dynamics of International Migration and Transnational Social Spaces*. Oxford University Press.
- Ferreira-Pereira, L. C., & Mourato Pinto, J. (2021). Soft Power in the European Union's Strategic Partnership Diplomacy: The Erasmus Plus Programme. In L. C. Ferreira-Pereira & M. Smith (Eds.), *The European Union's Strategic Partnerships: Global Diplomacy in a Contested World* (pp. 69–94). Springer International Publishing.
- Findlay, A. M. (2010). An Assessment of Supply and Demand-Side Theorizations of International Student Mobility. *International Migration*, 49(2), 162–190.
- Findlay, A. M., King, R., Smith, F. M., Geddes, A., & Skeldon, R. (2012). World Class? An Investigation of Globalisation, Difference and International Student Mobility. *Transactions of the Institute of British Geographers*, 37(1), 118–131.
- Findlay, A. M., & Li, F. L. N. (1998). A Migration Channels Approach to the Study of Professionals Moving to and from Hong Kong. *International Migration Review*, 32(3), 682–703.

- Fiorio, L., Zagheni, E., Abel, G. J., Hill, J., Pestre, G., Letouzé, E., & Cai, J. (2021). Analyzing the Effect of Time in Migration Measurement Using Georeferenced Digital Trace Data. *Demography*, 58(1), 51–74.
- Fitter, J., Raggl, A. K., & Ramskogler, P. (2024). *The (De)Globalization of Migration: Has the Polycrisis Period Changed the Patterns of Global Migration?* (Issue Q3/2024-3). Oesterreichische Nationalbank (Austrian Central Bank).
- FitzGerald, D. S. (2022). The Sociology of International Migration. In C. B. Brettell & J. F. Hollifield (Eds.), *Migration Theory: Talking Across Disciplines* (4th ed., pp. 160–193). Routledge.
- Fransen, S., & de Haas, H. (2022). Trends and Patterns of Global Refugee Migration. *Population and Development Review*, 48(1), 97–128.
- Fumasoli, T., & Rossi, F. (2021). The Role of Higher Education Institutions in Transnational Networks for Teaching and Learning Innovation: The Case of the Erasmus+ Programme. *European Journal of Education*, 56(2), 200–218.
- Fussell, E. (2010). The Cumulative Causation of International Migration in Latin America. *The ANNALS of the American Academy of Political and Social Science*, 630(1), 162–177.
- Fussell, E., & Massey, D. S. (2004). The Limits to Cumulative Causation: International Migration from Mexican Urban Areas. *Demography*, 41(1), 151–171.
- Gadár, L., Kosztyán, Z. T., Telcs, A., & Abonyi, J. (2022). Cooperation Patterns in the ERASMUS Student Exchange Network: An Empirical Study. *Applied Network Science*, 7(1), 74.
- Garip, F. (2012). Discovering Diverse Mechanisms of Migration: The Mexico–US Stream 1970–2000. *Population and Development Review*, 38(3), 393–433.
- Garip, F. (2019). *On the Move: Changing Mechanisms of Mexico–U.S. Migration*. Princeton University Press.
- Garip, F., & Asad, A. L. (2016). Network Effects in Mexico–U.S. Migration: Disentangling the Underlying Social Mechanisms. *American Behavioral Scientist*, 60(10), 1168–1193.
- Gold, S. J. (2005). Migrant Networks: A Summary and Critique of Relational Approaches to International Migration. In M. Romero & E. Margolis (Eds.), *The Blackwell Companion to Social Inequalities* (pp. 257–285). Wiley.

- González-Leonardo, M., López-Gay, A., Newsham, N., Recaño, J., & Rowe, F. (2022). Understanding Patterns of Internal Migration During the COVID-19 Pandemic in Spain. *Population, Space and Place*, 28(6), e2578.
- Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*, 91(3), 481–510.
- Gümüş, S., Gök, E., & Esen, M. (2020). A Review of Research on International Student Mobility: Science Mapping the Existing Knowledge Base. *Journal of Studies in International Education*, 24(5), 495–517.
- Ha, J. T., DeWaard, J., Abel, G. J., Tsuchiya, K., Pinchoff, J., Levesque, C., & Price, K. (2025). The Globalization of International Migration? A Conceptual and Data-Driven Synthesis. *Population and Development Review*, 51(1), 421–448.
- Haug, S. (2008). Migration Networks and Migration Decision-Making. *Journal of Ethnic and Migration Studies*, 34(4), 585–605.
- Hedström, P. (2005). *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge University Press.
- Hedström, P., & Bearman, P. (2009). What is Analytical Sociology All About? An Introductory Essay. In P. Hedström & P. Bearman (Eds.), *The Oxford Handbook of Analytical Sociology*. Oxford University Press.
- Hedström, P., & Ylikoski, P. (2010). Causal Mechanisms in the Social Sciences. *Annual Review of Sociology*, 36(1), 49–67.
- Heering, L., van der Erf, R. F., & Van Wissen, L. (2004). The Role of Family Networks and Migration Culture in the Continuation of Moroccan Emigration: a Gender Perspective. *Journal of Ethnic and Migration Studies*, 30(2), 323–337.
- Héran, F. (2022). Demography and Migration. In C. B. Brettell & J. F. Hollifield (Eds.), *Migration Theory: Talking Across Disciplines* (4th ed., pp. 78–129). Routledge.
- Hill, M. O. (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, 54(2), 427–432.
- Hoffman, M., & Block, P. (2026). Models for Emergent Structures in Mobility: Specification and Individual-Level Interpretation. *SocArXiv*.

- Hoffmann, R., Šedová, B., & Vinke, K. (2021). Improving the Evidence Base: A Methodological Review of the Quantitative Climate Migration Literature. *Global Environmental Change*, 71, 102367.
- Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., Margetts, H., Mullainathan, S., Salganik, M. J., Vazire, S., Vespignani, A., & Yarkoni, T. (2021). Integrating Explanation and Prediction in Computational Social Science. *Nature*, 595(7866), 181–188.
- Hou, C., & Du, D. (2022). The Changing Patterns of International Student Mobility: A Network Perspective. *Journal of Ethnic and Migration Studies*, 48(1), 248–272.
- Huang, P., & Butts, C. T. (2024). California Exodus? A Network Model of Population Redistribution in the United States. *The Journal of Mathematical Sociology*, 48(3), 311–339.
- Huang, P., & Butts, C. T. (2023). Rooted America: Immobility and Segregation of the Intercounty Migration Network. *American Sociological Review*, 88(6), 1031–1065.
- IMEM. (2012). Integrated Modelling of European Migration. Retrieved November 6, 2024, from www.imem.cpc.ac.uk
- Indlekofer, N., & Brandes, U. (2013). Relative Importance of Effects in Stochastic Actor-Oriented Models. *Network Science*, 1(3), 278–304.
- Jain, P. C. (2016). In Search of El Dorado: Indian Labour Migration to Gulf Countries. In P. C. Jain & G. Z. Oommen (Eds.), *South Asian Migration to Gulf Countries: History, Policies, Development* (pp. 136–152). Routledge.
- Jakobson, M.-L., Järvinen-Alenius, P., Pitkänen, P., Ruutsoo, R., Keski-Hirvelä, E., & Kalev, L. (2012). The Emergence of Estonian–Finnish Transnational Space. In P. Pitkänen, A. İçduygu & D. Sert (Eds.), *Migration and Transformation: Multi-Level Analysis of Migrant Transnationalism* (pp. 159–205). Springer Netherlands.
- Jayadeva, S. (2024). ‘Study-Abroad Influencers’ and Insider Knowledge: How New Forms of Study-Abroad Expertise on Social Media Mediate Student Mobility from India to Germany. *Mobilities*, 19(1), 1–17.
- Jayadeva, S. (2020). Keep Calm and Apply to Germany: How Online Communities Mediate Transnational Student Mobility from India to Germany. *Journal of Ethnic and Migration Studies*, 46(11), 2240–2257.

- Jokila, S., Kallo, J., & Mikkilä-Erdmann, M. (2019). From Crisis to Opportunities: Justifying and Persuading National Policy for International Student Recruitment. *European Journal of Higher Education*, 9(4), 393–411.
- Jost, L. (2006). Entropy and Diversity. *Oikos*, 113(2), 363–375.
- Jost, L. (2007). Partitioning Diversity into Independent Alpha and Beta Components. *Ecology*, 88(10), 2427–2439.
- Kandel, W., & Massey, D. S. (2002). The Culture of Mexican Migration: A Theoretical and Empirical Analysis. *Social Forces*, 80(3), 981–1004.
- Kashyap, R., Rinderknecht, R. G., Akbaritabar, A., Alburez-Gutierrez, D., Grow, A., Kim, J., Leasure, D. R., Lohmann, S., Negraia, D. V., Perrotta, D., Rampazzo, F., Tsai, C.-J., Verhagen, M. D., Zagheni, E., & Zhao, X. (2023). Digital and Computational Demography. In J. Skopek (Ed.), *Research Handbook on Digital Sociology* (pp. 47–85). Edward Elgar Publishing.
- Kashyap, R., & Villavicencio, F. (2016). The Dynamics of Son Preference, Technology Diffusion, and Fertility Decline Underlying Distorted Sex Ratios at Birth: A Simulation Approach. *Demography*, 53(5), 1261–1281.
- Kashyap, R., & Zagheni, E. (2023). Leveraging Digital and Computational Demography for Policy Insights. In E. Bertoni, M. Fontana, L. Gabrielli, S. Signorelli & M. Vespe (Eds.), *Handbook of Computational Social Science for Policy* (pp. 327–344). Springer International Publishing.
- Khan, M. A., & Hossain, M. I. (2017). The Emerging Phenomenon of Post-Globalized, South–South Migration: In Search of A Theoretical Framework. In P. Short, M. Hossain & M. A. Khan (Eds.), *South–South Migration: Emerging Patterns, Opportunities and Risks* (1st ed., pp. 11–33). Routledge.
- King, G., Tomz, M., & Wittenberg, J. (2000). Making the Most of Statistical Analyses: Improving Interpretation and Presentation. *American Journal of Political Science*, 44(2), 341–355.
- King, R. (1993). Recent Immigration to Italy: Character, Causes and Consequences. *GeoJournal*, 30(3), 283–292.
- King, R. (2025). Return Migration. In L. Oso, N. Ribas-Mateos & M. Moralli (Eds.), *Elgar Encyclopedia of Global Migration* (pp. 490–492). Edward Elgar Publishing.

- Kirk, D. (1960). Some Reflections on American Demography in the Nineteen Sixties. *Population Index*, 26(4), 305–310.
- Klabunde, A., & Willekens, F. (2016). Decision-Making in Agent-Based Models of Migration: State of the Art and Challenges. *European Journal of Population*, 32, 73–97.
- Klein, J. D., Weber, I., & Zagheni, E. (2024). Stop, in the Name of COVID! Using Social Media Data to Estimate the Effects of COVID-19-Related Travel Restrictions on Migration. *Demography*, 61(2), 493–511.
- Kniveton, D., Smith, C., & Wood, S. (2011). Agent-Based Model Simulations of Future Changes in Migration Flows for Burkina Faso. *Global Environmental Change*, 21, S34–S40.
- Kofman, E., Buhr, F., & Fonseca, M. L. (2022). Family Migration. In P. Scholten (Ed.), *Introduction to Migration Studies* (pp. 137–149). Springer International Publishing.
- Kondakci, Y., Bedenlier, S., & Zawacki-Richter, O. (2018). Social Network Analysis of International Student Mobility: Uncovering the Rise of Regional Hubs. *Higher Education*, 75(3), 517–535.
- Krivitsky, P. N. (2012). Exponential-Family Random Graph Models for Valued Networks. *Electronic Journal of Statistics*, 6, 1100–1128.
- Krzaklewska, E. (2008). Why Study Abroad? An Analysis of Erasmus Students' Motivations. In M. Byram & F. Dervin (Eds.), *Students, Staff and Academic Mobility in Higher Education* (1st ed., pp. 82–98). Cambridge Scholars Press.
- Kupiszewska, D., & Wiśniowski, A. (2009). *Availability of Statistical Data on Migration and Migrant Population and Potential Supplementary Sources for Data Estimation* (MIMOSA Project Deliverable 9.1A). Netherlands Interdisciplinary Demographic Institute (NIDI-KNAW) and University of Groningen. The Hague, Netherlands.
- Lazer, D. M. J., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational Social Science. *Science*, 323(5915), 721–723.
- Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational Social Science: Obstacles and Opportunities. *Science*, 369(6507), 1060–1062.

- Leal, D. F. (2021). Network Inequalities and International Migration in the Americas. *American Journal of Sociology*, 126(5), 1067–1126.
- Leal, D. F., & Harder, N. L. (2021). Global Dynamics of International Migration Systems Across South–South, North–North, and North–South flows, 1990–2015. *Applied Network Science*, 6(1).
- Leal, D. F., & Harder, N. L. (2023). Migration Networks and the Intensity of Global Migration Flows, 1990–2015. *Journal of Ethnic and Migration Studies*, 49(2), 445–464.
- Leinster, T. (2021). *Entropy and Diversity: The Axiomatic Approach* (1st ed.). Cambridge University Press.
- Lerpold, L., Sjöberg, Ö., & Wennberg, K. (2023). Migration, Integration, and the Pandemic. In L. Lerpold, Ö. Sjöberg & K. Wennberg (Eds.), *Migration and Integration in a Post-Pandemic World* (pp. 1–28). Springer International Publishing.
- Lesjak, M., Juvan, E., Ineson, E. M., Yap, M. H. T., & Axelsson, E. P. (2015). Erasmus Student Motivation: Why and Where to Go? *Higher Education*, 70(5), 845–865.
- Levy, N., Pisarevskaya, A., & Scholten, P. (2020). Between Fragmentation and Institutionalisation: The Rise of Migration Studies as a Research Field. *Comparative Migration Studies*, 8(1), 24.
- Lewer, J. J., & Van den Berg, H. (2008). A Gravity Model of Immigration. *Economics Letters*, 99(1), 164–167.
- Liang, Z., Chunyu, M. D., Zhuang, G., & Ye, W. (2008). Cumulative Causation, Market Transition, and Emigration from China. *American Journal of Sociology*, 114(3), 706–737.
- Lindstrom, D. P., & López Ramírez, A. (2010). Pioneers and Followers: Migrant Selectivity and the Development of U.S. Migration Streams in Latin America. *The ANNALS of the American Academy of Political and Social Science*, 630(1), 53–77.
- Liu, M.-M. (2013). Migrant Networks and International Migration: Testing Weak Ties. *Demography*, 50(4), 1243–1277.
- Lomax, N. (2022). Internal Migration: Census Data as a Gold Standard for Insight and Visualisation. *Journal of Maps*, 18(3), 543–550.
- Lospinoso, J., & Snijders, T. A. B. (2019). Goodness of Fit for Stochastic Actor-Oriented Models. *Methodological Innovations*, 12(3).

- Lusher, D., Koskinen, J., & Robins, G. (2012). *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press.
- Mabogunje, A. L. (1970). Systems Approach to a Theory of Rural-Urban Migration. *Geographical Analysis*, 2(1), 1–18.
- MacDonald, J. S., & MacDonald, L. D. (1964). Chain Migration, Ethnic Neighborhood Formation and Social Networks. *The Milbank Memorial Fund Quarterly*, 42(1), 82–97.
- Maciejewski, M. (2024). *The Maastricht and Amsterdam Treaties* [Fact Sheets on the European Union]. Retrieved July 6, 2025, from www.europarl.europa.eu/factsheets/en/sheet/3/the-maastricht-and-amsterdam-treaties
- Margetts, H., & Dorobantu, C. (2023). Computational Social Science for Public Policy. In E. Bertoni, M. Fontana, L. Gabrielli, S. Signorelli & M. Vespe (Eds.), *Handbook of Computational Social Science for Policy* (pp. 3–18). Springer International Publishing.
- Maringe, F., & de Wit, H. (2016). Global Higher Education Partnerships: Equity and Epistemic Concerns with Distribution and Flows of Intellectual Capital. In J. E. Côté & A. Furlong (Eds.), *Routledge Handbook of the Sociology of Higher Education*. Routledge.
- Martin, P. (2022). Economic Aspects of Migration. In C. B. Brettell & J. F. Hollifield (Eds.), *Migration Theory: Talking Across Disciplines* (4th ed., pp. 130–159). Routledge.
- Marzocchi, O. (2024). *Free Movement of Persons* [Fact Sheets on the European Union]. Retrieved July 6, 2025, from www.europarl.europa.eu/factsheets/en/sheet/147/free-movement-of-persons
- Massey, D. S. (1990). Social Structure, Household Strategies, and the Cumulative Causation of Migration. *Population Index*, 56(1), 3–26.
- Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1993). Theories of International Migration: A Review and Appraisal. *Population and Development Review*, 19(3), 431–466.
- Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1999). *Worlds in Motion: Understanding International Migration at the End of the Millennium*. Clarendon Press.

- Massey, D. S., & Espinosa, K. E. (1997). What's Driving Mexico–U.S. Migration? A Theoretical, Empirical, and Policy Analysis. *American Journal of Sociology*, 102(4), 939–999.
- Mata-Codesal, D., & Schmidt, K. (2020, February 14). The Mexico–US Migration Corridor. In T. Bastia & R. Skeldon (Eds.), *Routledge Handbook of Migration and Development* (1st ed., pp. 479–486). Routledge.
- Matthews, S. A., Stiberman, L., Raymer, J., Yang, T.-C., Gayawan, E., Saita, S., Tun, S. T. T., Parker, D. M., Balk, D., Leyk, S., Montgomery, M., Curtis, K. J., & Wong, D. W. S. (2021). Looking Back, Looking Forward: Progress and Prospect for Spatial Demography. *Spatial Demography*, 9(1), 1–29.
- Mayda, A. M. (2010). International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows. *Journal of Population Economics*, 23(4), 1249–1274.
- McMillan, C. (2024). Connecting the Dots: Using Social Network Analysis to Disentangle the Factors Driving International Migration. *International Migration*, imig.13234.
- Mihut, G., Cullinan, J., Flannery, D., Palcic, D., Souto-Otero, M., Wiers-Jenssen, J., Anabo, I. F., Kasza, G., Perez-Encinas, A., & Bin Qushem, U. (2025). International Student Mobility, Covid-19, and the Labour Market: A Scoping Review. *Comparative Migration Studies*, 13(11).
- Mood, C. (2010). Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review*, 26(1), 67–82.
- Muranaka, A. (2022). Beyond Blue Ocean? The Roles of Intermediaries in the Cross-Border Labour Market Between Japan and Vietnam. *Global Networks*, 22(3), 514–529.
- Murphy-Lejeune, E. (2002). *Student Mobility and Narrative in Europe: The New Strangers*. Routledge.
- Natter, K., & Welfens, N. (2024). Why Has Migration Research So Little Impact? Examining Knowledge Practices in Migration Policy Making and Migration Studies. *International Migration Review*, 58(4), 1669–1700.
- Office for National Statistics. (2020). International Passenger Survey 4.03, Country of Birth by Country of Last or Next Residence (Discontinued after 2019). Retrieved November 6, 2024, from

- www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/datasets/ipscountryofbirthbycountryoflastornextresidence
- Ovchinnikova, E., Van Mol, C., & Jones, E. (2023). The Role of Language Proximity in Shaping International Student Mobility Flows. *Globalisation, Societies and Education*, 21(4), 563–574.
- Özden, Ç., Parsons, C. R., Schiff, M., & Walmsley, T. L. (2011). Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960–2000. *The World Bank Economic Review*, 25(1), 12–56.
- Palloni, A., Massey, D. S., Ceballos, M., Espinosa, K., & Spittel, M. (2001). Social Capital and International Migration: A Test Using Information on Family Networks. *American Journal of Sociology*, 106(5), 1262–1298.
- Papatsiba, V., & Marginson, S. (2025). *Brexit, EU Students and UK Higher Education: Broken Bridges* (1st ed). Bloomsbury Academic.
- Paredes-Orozco, G. (2019). The Limits to Cumulative Causation Revisited: Urban-Origin Mexico–US Migration in an Era of Increased Immigration Restrictions. *Demographic Research*, 41, 815–846.
- Perez-Encinas, A., Rodriguez-Pomeda, J., & de Wit, H. (2020). Factors Influencing Student Mobility: A Comparative European Study. *Studies in Higher Education*, 46(12), 2528–2541.
- Perkins, R., & Neumayer, E. (2014). Geographies of Educational Mobilities: Exploring the Uneven Flows of International Students. *The Geographical Journal*, 180(3), 246–259.
- Pfaffermayr, M. (2023). Cross-Sectional Gravity Models, PPML Estimation, and the Bias Correction of the Two-Way Cluster-Robust Standard Errors. *Oxford Bulletin of Economics and Statistics*, 85(5), 1111–1134.
- Plane, D. A., & Mulligan, G. F. (1997). Measuring Spatial Focusing in a Migration System. *Demography*, 34(2), 251–262.
- Plotnikova, M., & Ulceluse, M. (2022). Inequality as a Driver of Migration: A Social Network Analysis. *Population, Space and Place*, 28(8).

- Poot, J., Alimi, O., Cameron, M. P., & Maré, D. C. (2016). The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science. *IZA Discussion Paper No. 10329*.
- Price, C. A. (1963). *Southern Europeans in Australia*. Oxford University Press.
- Price, M. (2022). Geographical Theories of Migration. In C. B. Brettell & J. F. Hollifield (Eds.), *Migration Theory: Talking Across Disciplines* (4th ed., pp. 232–268). Routledge.
- Przytuła, S. (2015). Migrants, Assigned Expatriates (AE) and Self-Initiated Expatriates (SIE)—Differentiation of Terms and Literature-Based Research Review. *Journal of Intercultural Management*, 7(2), 89–111.
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved December 26, 2025, from www.R-project.org
- Raftery, A. E. (2000). Statistics in Sociology, 1950–2000. *Journal of the American Statistical Association*, 95(450), 654–661.
- Ramos, R. (2016). Gravity Models: A Tool for Migration Analysis. *IZA World of Labor*, 239.
- Ramos, R., & Suriñach, J. (2017). A Gravity Model of Migration Between the ENC and the EU. *Tijdschrift voor Economische en Sociale Geografie*, 108(1), 21–35.
- Ravenstein, E. G. (1885). The Laws of Migration I. *Journal of the Statistical Society of London*, 48(2), 167–235.
- Ravenstein, E. G. (1889). The Laws of Migration II. *Journal of the Royal Statistical Society*, 52(2), 214–301.
- Raymer, J. (2016). Migration in Europe. In M. J. White (Ed.), *International Handbook of Migration and Population Distribution* (pp. 371–388, Vol. 6). Springer Netherlands.
- Raymer, J., & Rogers, A. (2007). Using Age and Spatial Flow Structures in the Indirect Estimation of Migration Streams. *Demography*, 44(2), 199–223.
- Raymer, J., Willekens, F., & Rogers, A. (2019). Spatial Demography: A Unifying Core and Agenda for Further Research. *Population, Space and Place*, 25(4), e2179.
- Raymer, J., Wiśniowski, A., Forster, J. J., Smith, P. W., & Bijak, J. (2013). Integrated Modeling of European Migration. *Journal of the American Statistical Association*, 108(503), 801–819.

- Recchi, E. (2015). *Mobile Europe: The Theory and Practice of Free Movement in the EU*. Palgrave Macmillan UK.
- Restaino, M., Vitale, M. P., & Primerano, I. (2020). Analysing International Student Mobility Flows in Higher Education: A Comparative Study on European Countries. *Social Indicators Research*, 149(3), 947–965.
- Riosmena, F. (2024). Worlds in Motion *Redux*? Expanding Migration Theories and Their Interconnections. *Population and Development Review*, 50(3), 677–726.
- Rodríguez González, C., Mesanza, R. B., & Mariel, P. (2011). The Determinants of International Student Mobility Flows: An Empirical Study on the Erasmus Programme. *Higher Education*, 62(4), 413–430.
- Rogers, A., & Raymer, J. (1998). The Spatial Focus of US Interstate Migration Flows. *International Journal of Population Geography*, 4(1), 63–80.
- Rogers, A., & Sweeney, S. (1998). Measuring the Spatial Focus of Migration Patterns. *The Professional Geographer*, 50(2), 232–242.
- Santos Silva, J. M. C., & Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, 88(4), 641–658.
- Santos Silva, J. M. C., & Tenreyro, S. (2022). The Log of Gravity at 15. *Portuguese Economic Journal*, 21(3), 423–437.
- Schmid, M. (2011). The Logic of Mechanistic Explanations in the Social Sciences. In P. Demeulenaere (Ed.), *Analytical Sociology and Social Mechanisms* (pp. 136–153). Cambridge University Press.
- Scholten, P., Pisarevskaya, A., & Levy, N. (2022). An Introduction to Migration Studies: The Rise and Coming of Age of a Research Field. In P. Scholten (Ed.), *Introduction to Migration Studies* (pp. 3–24). Springer International Publishing.
- Seven, Ü. (2022). Armed Conflict, Violence, and the Decision to Migrate: Explaining the Determinants of Displacement in Syria. *Migration and Development*, 11(3), 1029–1045.
- Shields, R. (2013). Globalization and International Student Mobility: A Network Analysis. *Comparative Education Review*, 57(4), 609–636.
- Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, 25(3), 289–310.

- Sin, C., Antonowicz, D., & Wiers-Jenssen, J. (2021). Attracting International Students to Semi-Peripheral Countries: A Comparative Study of Norway, Poland and Portugal. *Higher Education Policy*, 34(1), 297–320.
- Snijders, T. A. B. (2002). Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. *Journal of Social Structure*, 3(2), 1–40.
- Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New Specifications for Exponential Random Graph Models. *Sociological Methodology*, 36(1), 99–153.
- Souto-Otero, M., Huisman, J., Beerkens, M., de Wit, H., & Vujić, S. (2013). Barriers to International Student Mobility. *Educational Researcher*, 42(2), 70–77.
- Stadtfeld, C., & Amati, V. (2021, December 17). Network Mechanisms and Network Models. In G. Manzo (Ed.), *Research Handbook on Analytical Sociology* (pp. 432–452). Edward Elgar Publishing.
- Stouffer, S. A. (1960). Intervening Opportunities and Competing Migrants. *Journal of Regional Science*, 2(1), 1–26.
- Sumption, M. (2019). Is Employer Sponsorship a Good Way to Manage Labour Migration? Implications for Post-Brexit Migration Policies. *National Institute Economic Review*, 248, R28–R39.
- Sumption, M. (2022). Shortages, High-Demand Occupations, and the Post-Brexit UK Immigration System. *Oxford Review of Economic Policy*, 38(1), 97–111.
- Teichler, U. (2019). Bologna and Student Mobility: A Fuzzy Relationship. *Innovation: The European Journal of Social Science Research*, 32(4), 429–449.
- Teichler, U. (2012). International Student Mobility and the Bologna Process. *Research in Comparative and International Education*, 7(1), 34–49.
- Teichler, U. (2017). Internationalisation Trends in Higher Education and the Changing Role of International Student Mobility. *Journal of International Mobility*, 5(1), 177–216.
- Teichler, U. (2004). Temporary Study Abroad: The Life of ERASMUS Students. *European Journal of Education*, 39(4), 395–408.
- Termote, M. (2002). The Explanatory Power of Migration Models. In R. Franck (Ed.), *The Explanatory Power of Models* (1st ed., pp. 165–180, Vol. 1). Springer Science & Business Media.

- Thomas, L. J., Christakis, N. A., & Garip, F. (2026). Social Networks and International Migration in Honduran Villages. *Proceedings of the National Academy of Sciences*, 123(1), e2505818122.
- Toma, S., & Vause, S. (2014). Gender Differences in the Role of Migrant Networks: Comparing Congolese and Senegalese Migration Flows. *International Migration Review*, 48(4), 972–997.
- Ton, M. J., De Moel, H., De Bruijn, J. A., Botzen, W. J., Karabiyik, H., Friedrich, M., & Aerts, J. C. (2024). The Impact of Natural Hazards on Migration in the United States and the Effect of Spatial Dependence. *Journal of Environmental Planning and Management*, 69(1), 28–46.
- Toomistu, T., Lauren, A., Annist, A., & Murakas, R. (2024). Determinants of Return Migration of Estonian Young Adults in Transnational Mobility. *Migration Studies*, 12(1), 42–67.
- Tranos, E., Gheasi, M., & Nijkamp, P. (2015). International Migration: A Global Complex Network. *Environment and Planning B: Planning and Design*, 42(1), 4–22.
- UNESCO Institute for Statistics. (2023). UIS Statistics: Total Inbound Internationally Mobile Students by Country of Origin, Both Sexes (Number). Retrieved May 18, 2023, from <http://data.uis.unesco.org>
- United Nations Department of Economic and Social Affairs. (2019). *International Migrant Stock 2019*. United Nations Department of Economic and Social Affairs/Population Division. New York, NY.
- United Nations Department of Economic and Social Affairs. (2015). *Trends in International Migrant Stock: The 2015 Revision*. United Nations Department of Economic and Social Affairs/Population Division. New York, NY.
- United Nations Statistics Division. (2024). *Standard Country or Area Codes for Statistical Use (M49)*. United Nations. Retrieved December 26, 2025, from <https://unstats.un.org/unsd/methodology/m49/>
- Van Bouwel, L., & Veugelers, R. (2013). The Determinants of Student Mobility in Europe: The Quality Dimension. *European Journal of Higher Education*, 3(2), 172–190.
- van der Erf, R. F. (2009). *Typology of Data and Feasibility Study* (MIMOSA Project Deliverable 9.1B). Netherlands Interdisciplinary Demographic Institute (NIDI-KNAW) and University of Groningen. The Hague, Netherlands.

- Van Hear, N., Bakewell, O., & Long, K. (2018). Push-Pull Plus: Reconsidering the Drivers of Migration. *Journal of Ethnic and Migration Studies*, 44(6), 927–944.
- Van Mol, C. (2014). *Intra-European Student Mobility in International Higher Education Circuits*. Palgrave Macmillan.
- Van Mol, C., & Timmerman, C. (2014). Should I Stay or Should I Go? An Analysis of the Determinants of Intra-European Student Mobility. *Population, Space and Place*, 20(5), 465–479.
- Vatta, A. (2017). The EU Migration Policy Between Europeanization and Re-Nationalization. In S. Baldin & M. Zago (Eds.), *Europe of Migrations: Policies, Legal Issues and Experiences* (1st ed., pp. 13–32). Edizioni Università di Trieste.
- Vögtle, E., & Windzio, M. (2022). The ‘Global South’ in the Transnational Student Mobility Network. Effects of Institutional Instability, Reputation of the Higher Education Systems, Post-Colonial Ties, and Culture. *Globalisation, Societies and Education*, 23(4), 903–921.
- Waldorf, B. (1995). Determinants of International Return Migration Intentions. *The Professional Geographer*, 47(2), 125–136.
- Wassink, J., & Massey, D. S. (2022). The New System of Mexican Migration: The Role of Entry Mode-Specific Human and Social Capital. *Demography*, 59(3), 1071–1092.
- Weber, T., & Van Mol, C. (2023). The Student Migration Transition: An Empirical Investigation into the Nexus Between Development and International Student Migration. *Comparative Migration Studies*, 11(1).
- Weidner, M., & Zylkin, T. (2021). Bias and Consistency in Three-Way Gravity Models. *Journal of International Economics*, 132, 103513.
- Windzio, M. (2018). The Network of Global Migration 1990–2013. *Social Networks*, 53, 20–29.
- Wittek, B. H. (2021). Where Are We Today in Understanding the Role of the Firm as a Driver of International Flows of Highly Skilled Migrants? Reviewing the Status of the Literature on Direct and Indirect Host Country Meso-Level Influences. *The International Journal of Human Resource Management*, 32(20), 4253–4286.
- World Bank. (2025a). GDP per Capita, PPP (Current International \$). Retrieved July 6, 2025, from <https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD>

- World Bank. (2025b). Population, Total. Retrieved July 6, 2025, from <https://data.worldbank.org/indicator/SP.POP.TOTL>
- World Bank. (2023). *World Development Report 2023: Migrants, Refugees, and Societies*.
- Xiang, B., & Lindquist, J. (2014). Migration Infrastructure. *International Migration Review*, 48(1), 122–148.
- Yang, P., & Yi, T. (2023). International Student Mobility in a Deglobalizing and Post-Pandemic World: Resilience, Reconfiguration, Renewal. *Journal of International Students*, 13(1), i–v.
- Yildiz, D., & Abel, G. J. (2021). Migration Stocks and Flows: Data Concepts, Availability and Comparability. In M. McAuliffe (Ed.), *Research Handbook on International Migration and Digital Technology* (pp. 29–41). Edward Elgar Publishing.
- Ylikoski, P. (2011). Social Mechanisms and Explanatory Relevance. In P. Demeulenaere (Ed.), *Analytical Sociology and Social Mechanisms* (pp. 154–172). Cambridge University Press.
- Zhao, C., & Qu, X. (2022). Social Networks and Internal Migration in China: A Spatial Autoregressive Model. *Review of Development Economics*, 26(2), 1132–1163.

A | Appendix to Chapter 3

A.1 Data Availability Statement

The data used in this chapter consist of novel bilateral migration flow estimates derived from publicly available migrant stock data released by the World Bank (Özden et al., 2011; World Bank, 2023). These stock data, harmonised to contemporary national borders and measured primarily by country of birth, are used to estimate country-to-country migration flows covering 195 countries over six ten-year periods from 1960–1970 to 2010–2020.

Migration flows were estimated with the *migest* R package (Abel, 2025) using the statistical software R (version 4.3.2; R Core Team, 2023), applying three demographic accounting methods: (i) pseudo-Bayesian estimation under a closed-population assumption, (ii) minimisation under a closed-population assumption, and (iii) minimisation under an open-population assumption. The pseudo-Bayesian estimates are used in the main analysis, while the two minimisation-based variants are used for robustness checks.

Because the bilateral flow data analysed in this chapter were constructed by the author rather than obtained as a precompiled third-party dataset, the estimated flow matrices are not currently available through an external data provider. Upon acceptance of the article version of this chapter, the full set of bilateral flow estimates will be released, along with all data-processing scripts and code used to estimate the flows, compute the diversity measures, and reproduce the empirical analyses, in a dedicated public online repository.

A.2 List of Countries Included in the Analysis

Table A.1: List of countries included in the analysis, with corresponding ISO 3166-1 codes (alpha-3), continent, and UN world region labels.

ISO3	Country name	Continent	Region
AFG	Afghanistan	Asia	Southern Asia
AGO	Angola	Africa	Middle & Southern Africa
ALB	Albania	Europe	Southern Europe
AND	Andorra	Europe	Southern Europe
ARE	United Arab Emirates	Asia	Central Asia & Western Asia
ARG	Argentina	Americas	Southern America
ARM	Armenia	Asia	Central Asia & Western Asia
ATG	Antigua and Barbuda	Americas	Central America & Caribbean
AUS	Australia	Oceania	Oceania
AUT	Austria	Europe	Western Europe
AZE	Azerbaijan	Asia	Central Asia & Western Asia
BDI	Burundi	Africa	Eastern Africa
BEL	Belgium	Europe	Western Europe
BEN	Benin	Africa	Western Africa
BFA	Burkina Faso	Africa	Western Africa
BGD	Bangladesh	Asia	Southern Asia
BGR	Bulgaria	Europe	Eastern Europe
BHR	Bahrain	Asia	Central Asia & Western Asia
BHS	Bahamas	Americas	Central America & Caribbean
BIH	Bosnia and Herzegovina	Europe	Southern Europe
BLR	Belarus	Europe	Eastern Europe
BLZ	Belize	Americas	Central America & Caribbean
BOL	Bolivia (Plurinational State of)	Americas	Southern America

Table A.1 continued from previous page

ISO3	Country name	Continent	Region
BRA	Brazil	Americas	Southern America
BRB	Barbados	Americas	Central America & Caribbean
BRN	Brunei Darussalam	Asia	South-Eastern Asia
BTN	Bhutan	Asia	Southern Asia
BWA	Botswana	Africa	Middle & Southern Africa
CAF	Central African Republic	Africa	Middle & Southern Africa
CAN	Canada	Americas	North America
CHE	Switzerland	Europe	Western Europe
CHL	Chile	Americas	Southern America
CHN	China (incl. Taiwan)	Asia	Eastern Asia
CIV	Côte d'Ivoire	Africa	Western Africa
CMR	Cameroon	Africa	Middle & Southern Africa
COD	Democratic Republic of the Congo	Africa	Middle & Southern Africa
COG	Congo	Africa	Middle & Southern Africa
COL	Colombia	Americas	Southern America
COM	Comoros	Africa	Eastern Africa
CPV	Cabo Verde	Africa	Western Africa
CRI	Costa Rica	Americas	Central America & Caribbean
CUB	Cuba	Americas	Central America & Caribbean
CYP	Cyprus	Asia	Central Asia & Western Asia
CZE	Czechia	Europe	Eastern Europe
DEU	Germany	Europe	Western Europe
DJI	Djibouti	Africa	Eastern Africa
DMA	Dominica	Americas	Central America & Caribbean
DNK	Denmark	Europe	Northern Europe
DOM	Dominican Republic	Americas	Central America & Caribbean
DZA	Algeria	Africa	Northern Africa

Table A.1 continued from previous page

ISO3	Country name	Continent	Region
ECU	Ecuador	Americas	Southern America
EGY	Egypt	Africa	Northern Africa
ERI	Eritrea	Africa	Eastern Africa
ESP	Spain	Europe	Southern Europe
EST	Estonia	Europe	Northern Europe
ETH	Ethiopia	Africa	Eastern Africa
FIN	Finland	Europe	Northern Europe
FJI	Fiji	Oceania	Oceania
FRA	France	Europe	Western Europe
FSM	Micronesia (Fed. States of)	Oceania	Oceania
GAB	Gabon	Africa	Middle & Southern Africa
GBR	United Kingdom	Europe	Northern Europe
GEO	Georgia	Asia	Central Asia & Western Asia
GHA	Ghana	Africa	Western Africa
GIN	Guinea	Africa	Western Africa
GMB	Gambia	Africa	Western Africa
GNB	Guinea-Bissau	Africa	Western Africa
GNQ	Equatorial Guinea	Africa	Middle & Southern Africa
GRC	Greece	Europe	Southern Europe
GRD	Grenada	Americas	Central America & Caribbean
GTM	Guatemala	Americas	Central America & Caribbean
GUY	Guyana	Americas	Southern America
HKG	China, Hong Kong SAR	Asia	Eastern Asia
HND	Honduras	Americas	Central America & Caribbean
HRV	Croatia	Europe	Southern Europe
HTI	Haiti	Americas	Central America & Caribbean
HUN	Hungary	Europe	Eastern Europe

Table A.1 continued from previous page

ISO3	Country name	Continent	Region
IDN	Indonesia	Asia	South-Eastern Asia
IND	India	Asia	Southern Asia
IRL	Ireland	Europe	Northern Europe
IRN	Iran (Islamic Republic of)	Asia	Southern Asia
IRQ	Iraq	Asia	Central Asia & Western Asia
ISL	Iceland	Europe	Northern Europe
ISR	Israel	Asia	Central Asia & Western Asia
ITA	Italy	Europe	Southern Europe
JAM	Jamaica	Americas	Central America & Caribbean
JOR	Jordan	Asia	Central Asia & Western Asia
JPN	Japan	Asia	Eastern Asia
KAZ	Kazakhstan	Asia	Central Asia & Western Asia
KEN	Kenya	Africa	Eastern Africa
KGZ	Kyrgyzstan	Asia	Central Asia & Western Asia
KHM	Cambodia	Asia	South-Eastern Asia
KIR	Kiribati	Oceania	Oceania
KNA	Saint Kitts and Nevis	Americas	Central America & Caribbean
KOR	Republic of Korea	Asia	Eastern Asia
KWT	Kuwait	Asia	Central Asia & Western Asia
LAO	Lao People's Democratic Republic	Asia	South-Eastern Asia
LBN	Lebanon	Asia	Central Asia & Western Asia
LBR	Liberia	Africa	Western Africa
LBY	Libya	Africa	Northern Africa
LCA	Saint Lucia	Americas	Central America & Caribbean
LIE	Liechtenstein	Europe	Western Europe
LKA	Sri Lanka	Asia	Southern Asia
LSO	Lesotho	Africa	Middle & Southern Africa

Table A.1 continued from previous page

ISO3	Country name	Continent	Region
LTU	Lithuania	Europe	Northern Europe
LUX	Luxembourg	Europe	Western Europe
LVA	Latvia	Europe	Northern Europe
MAC	China, Macao SAR	Asia	Eastern Asia
MAR	Morocco	Africa	Northern Africa
MCO	Monaco	Europe	Western Europe
MDA	Republic of Moldova	Europe	Eastern Europe
MDG	Madagascar	Africa	Eastern Africa
MDV	Maldives	Asia	Southern Asia
MEX	Mexico	Americas	Central America & Caribbean
MHL	Marshall Islands	Oceania	Oceania
MKD	North Macedonia	Europe	Southern Europe
MLI	Mali	Africa	Western Africa
MLT	Malta	Europe	Southern Europe
MMR	Myanmar	Asia	South-Eastern Asia
MNG	Mongolia	Asia	Eastern Asia
MOZ	Mozambique	Africa	Eastern Africa
MRT	Mauritania	Africa	Western Africa
MUS	Mauritius	Africa	Eastern Africa
MWI	Malawi	Africa	Eastern Africa
MYS	Malaysia	Asia	South-Eastern Asia
NAM	Namibia	Africa	Middle & Southern Africa
NER	Niger	Africa	Western Africa
NGA	Nigeria	Africa	Western Africa
NIC	Nicaragua	Americas	Central America & Caribbean
NIU	Niue	Oceania	Oceania
NLD	Netherlands	Europe	Western Europe

Table A.1 continued from previous page

ISO3	Country name	Continent	Region
NOR	Norway	Europe	Northern Europe
NPL	Nepal	Asia	Southern Asia
NRU	Nauru	Oceania	Oceania
NZL	New Zealand	Oceania	Oceania
OMN	Oman	Asia	Central Asia & Western Asia
PAK	Pakistan	Asia	Southern Asia
PAN	Panama	Americas	Central America & Caribbean
PER	Peru	Americas	Southern America
PHL	Philippines	Asia	South-Eastern Asia
PLW	Palau	Oceania	Oceania
PNG	Papua New Guinea	Oceania	Oceania
POL	Poland	Europe	Eastern Europe
PRK	Dem. People's Republic of Korea	Asia	Eastern Asia
PRT	Portugal	Europe	Southern Europe
PRY	Paraguay	Americas	Southern America
PSE	State of Palestine	Asia	Central Asia & Western Asia
QAT	Qatar	Asia	Central Asia & Western Asia
ROU	Romania	Europe	Eastern Europe
RUS	Russian Federation	Europe	Eastern Europe
RWA	Rwanda	Africa	Eastern Africa
SAU	Saudi Arabia	Asia	Central Asia & Western Asia
SCG	Serbia and Montenegro	Europe	Southern Europe
SEN	Senegal	Africa	Western Africa
SGP	Singapore	Asia	South-Eastern Asia
SLB	Solomon Islands	Oceania	Oceania
SLE	Sierra Leone	Africa	Western Africa
SLV	El Salvador	Americas	Central America & Caribbean

Table A.1 continued from previous page

ISO3	Country name	Continent	Region
SMR	San Marino	Europe	Southern Europe
SOM	Somalia	Africa	Eastern Africa
STP	Sao Tome and Principe	Africa	Middle & Southern Africa
SUD	Sudan (incl. South Sudan)	Africa	Northern Africa
SUR	Suriname	Americas	Southern America
SVK	Slovakia	Europe	Eastern Europe
SVN	Slovenia	Europe	Southern Europe
SWE	Sweden	Europe	Northern Europe
SWZ	Eswatini	Africa	Middle & Southern Africa
SYC	Seychelles	Africa	Eastern Africa
SYR	Syrian Arab Republic	Asia	Central Asia & Western Asia
TCD	Chad	Africa	Middle & Southern Africa
TGO	Togo	Africa	Western Africa
THA	Thailand	Asia	South-Eastern Asia
TJK	Tajikistan	Asia	Central Asia & Western Asia
TKM	Turkmenistan	Asia	Central Asia & Western Asia
TLS	Timor-Leste	Asia	South-Eastern Asia
TON	Tonga	Oceania	Oceania
TTO	Trinidad and Tobago	Americas	Central America & Caribbean
TUN	Tunisia	Africa	Northern Africa
TUR	Türkiye	Asia	Central Asia & Western Asia
TUV	Tuvalu	Oceania	Oceania
TZA	United Republic of Tanzania	Africa	Eastern Africa
UGA	Uganda	Africa	Eastern Africa
UKR	Ukraine	Europe	Eastern Europe
URY	Uruguay	Americas	Southern America
USA	United States of America	Americas	North America

Table A.1 continued from previous page

ISO3	Country name	Continent	Region
UZB	Uzbekistan	Asia	Central Asia & Western Asia
VCT	Saint Vincent and the Grenadines	Americas	Central America & Caribbean
VEN	Venezuela (Bolivarian Republic of)	Americas	Southern America
VNM	Viet Nam	Asia	South-Eastern Asia
VUT	Vanuatu	Oceania	Oceania
WSM	Samoa	Oceania	Oceania
YEM	Yemen	Asia	Central Asia & Western Asia
ZAF	South Africa	Africa	Middle & Southern Africa
ZMB	Zambia	Africa	Eastern Africa
ZWE	Zimbabwe	Africa	Eastern Africa

A.3 Distribution of Corridor Volumes, 1960–2020, Pseudo-Bayesian Flow Estimation

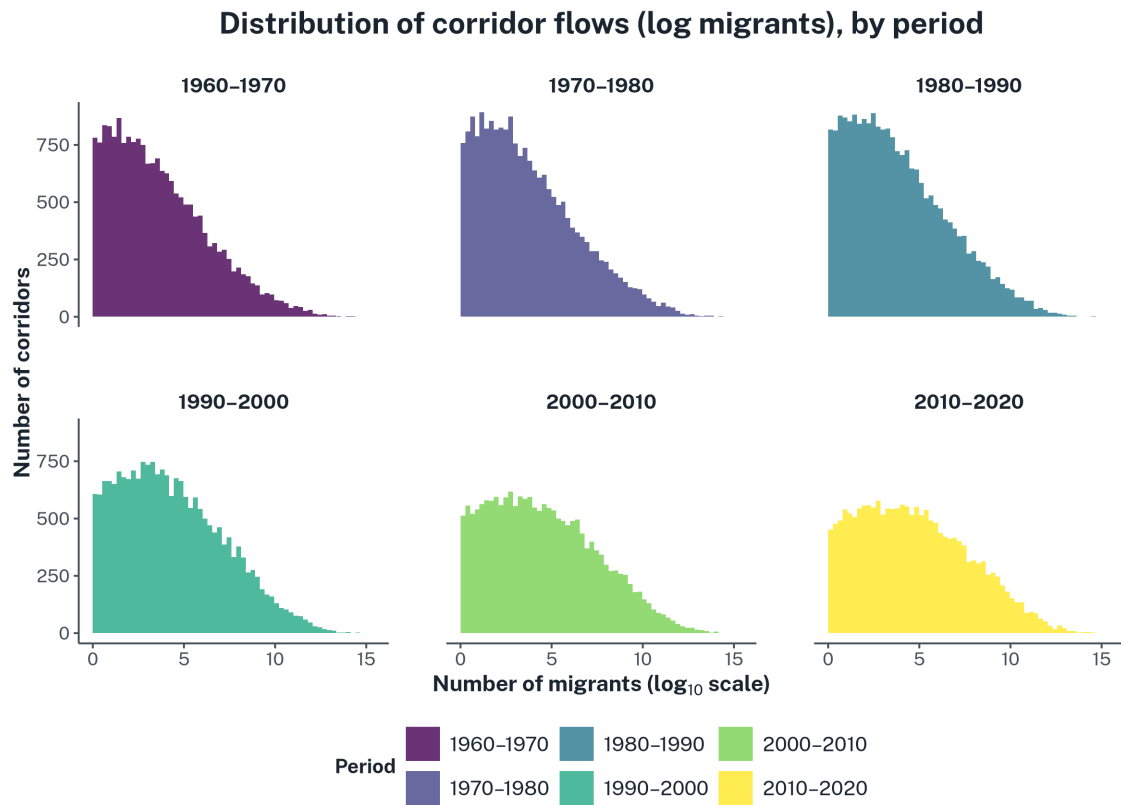


Figure A.1: Distribution of corridor volumes over time, 1960–2020, using pseudo-Bayesian flow estimates.

A.4 Choropleth World Maps of Total Inflows and Outflows, Pseudo-Bayesian Flow Estimation

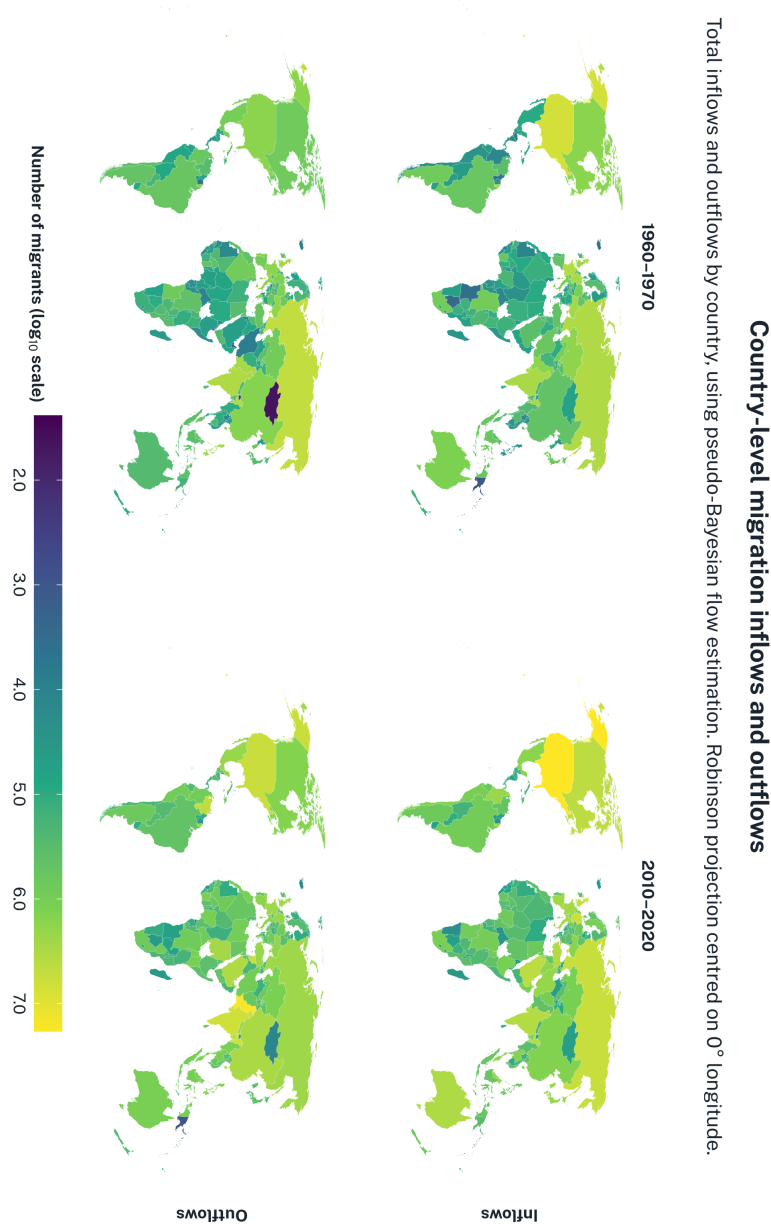


Figure A.2: Choropleth maps of country-level migration inflows and outflows in the first (1960–1970) and last (2010–2020) periods under study, using pseudo-Bayesian flow estimates. Values are shown on a log₁₀ scale. Maps use the Robinson projection centred on 0° longitude.

A.5 Chord Diagrams of Directional Migration Flows by World Region, Pseudo-Bayesian Flow Estimation

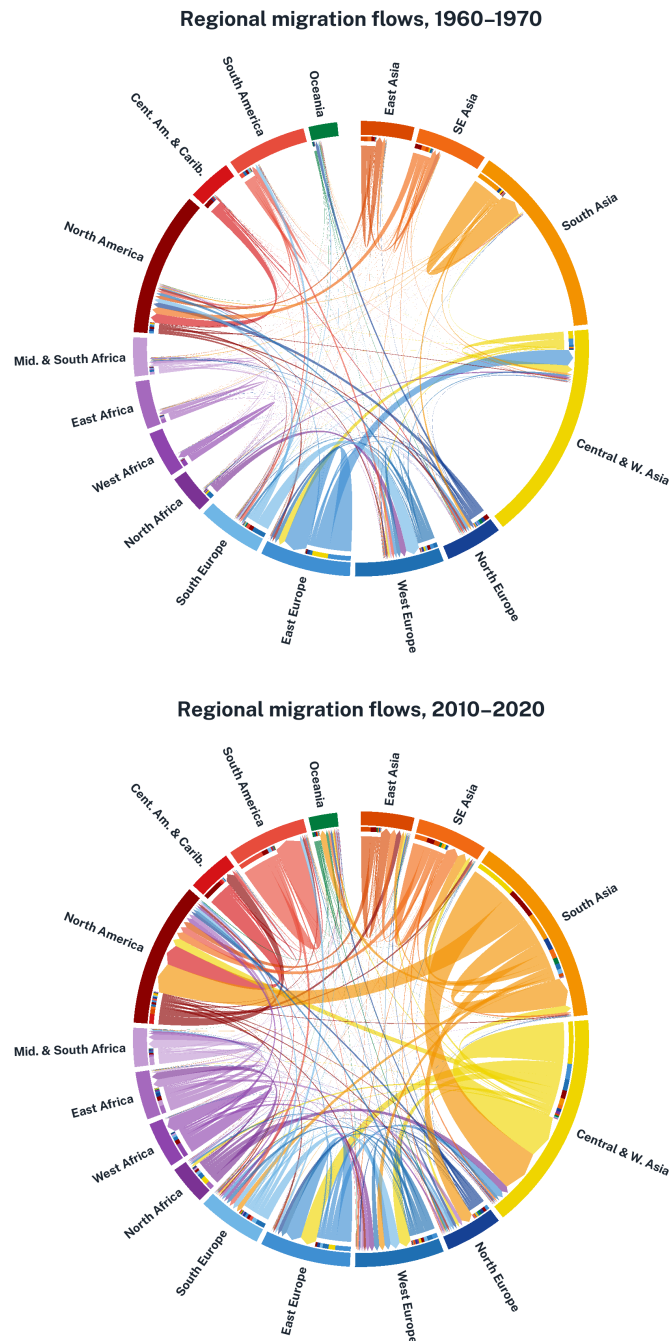


Figure A.3: Chord diagrams of region-to-region migration flows in the first (1960–1970) and last (2010–2020) periods under study, using pseudo-Bayesian flow estimates.

A.6 Hill Diversity of Migration Inflows and Outflows for All Countries, Pseudo-Bayesian Flow Estimation

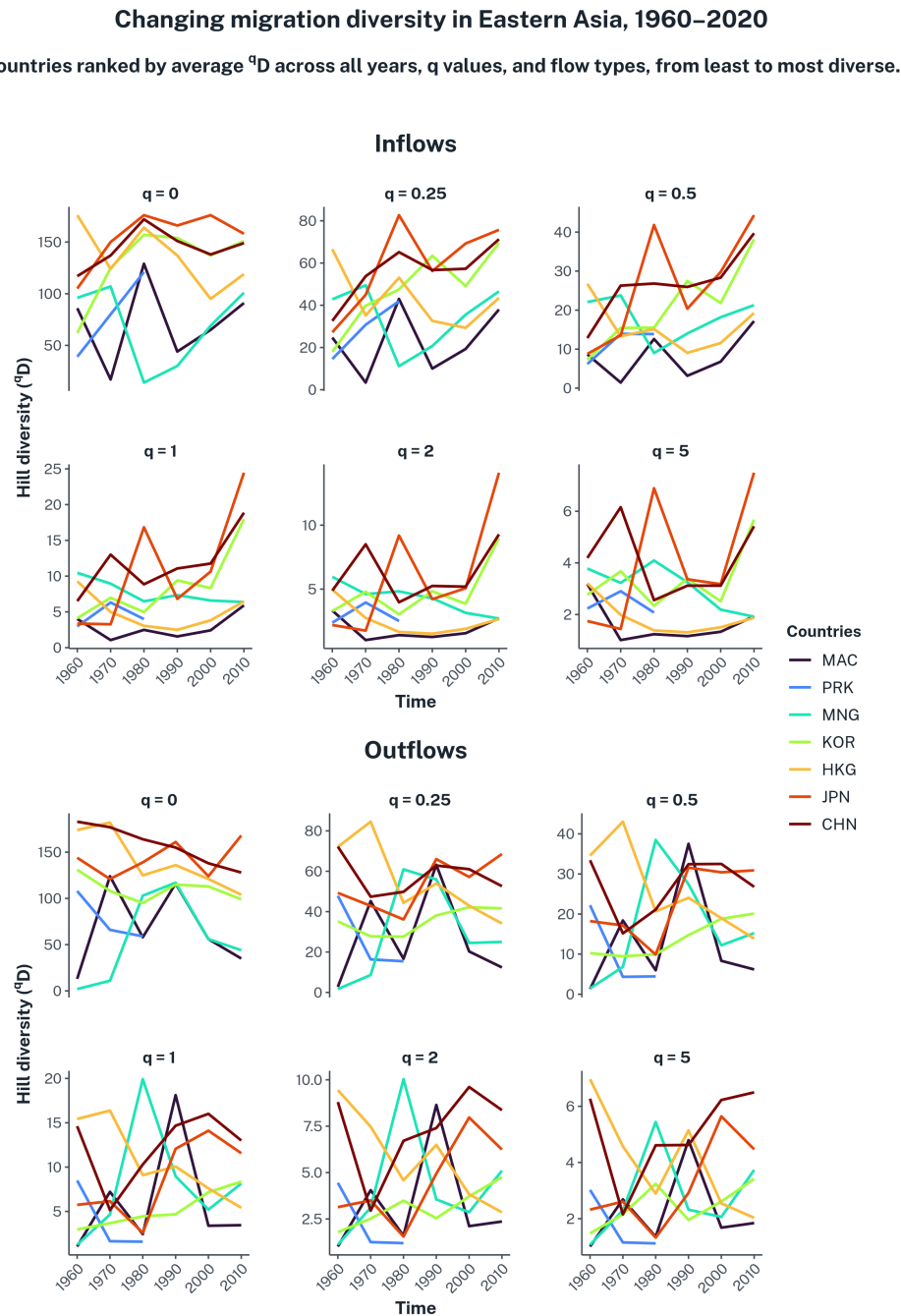


Figure A.4: Migration diversity in inflows and outflows for East Asian countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in South-Eastern Asia, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

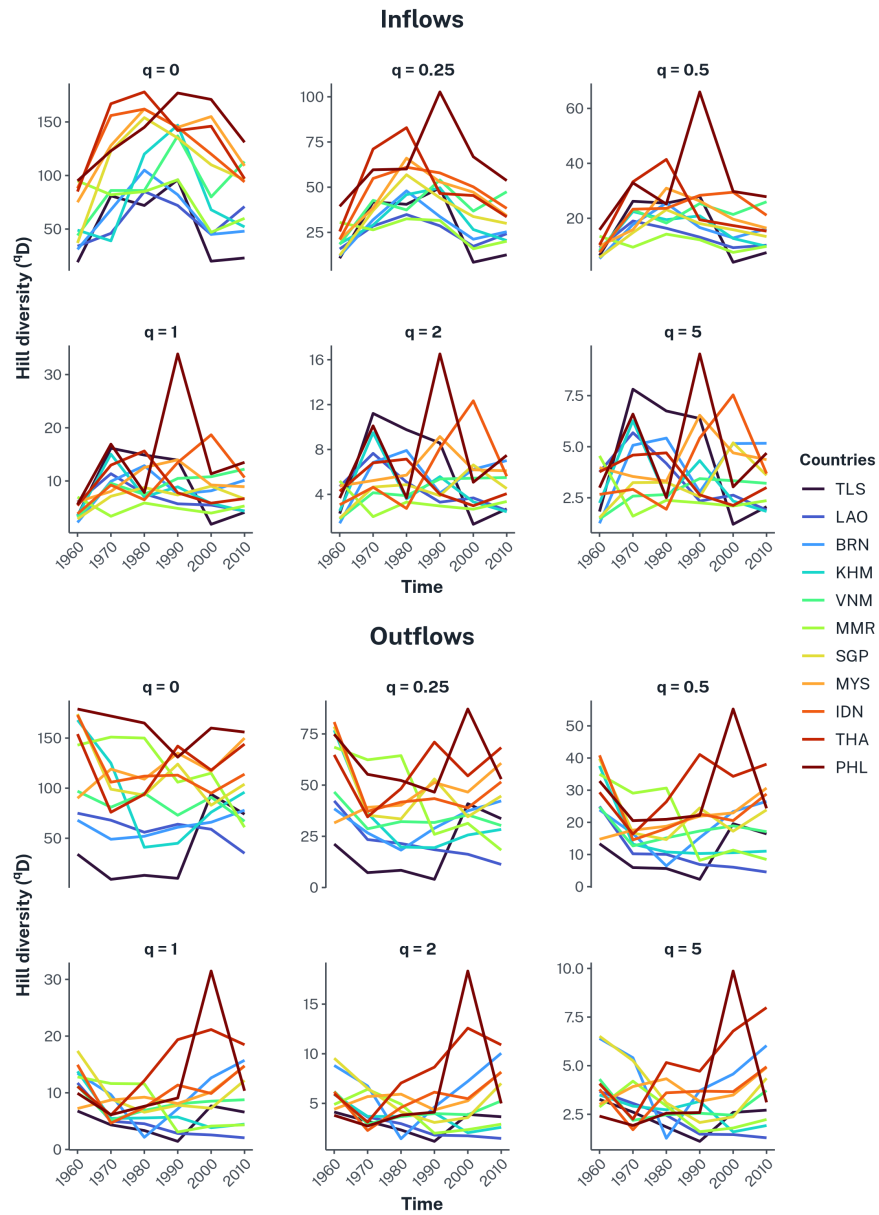


Figure A.5: Migration diversity in inflows and outflows for South-East Asian countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in Southern Asia, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

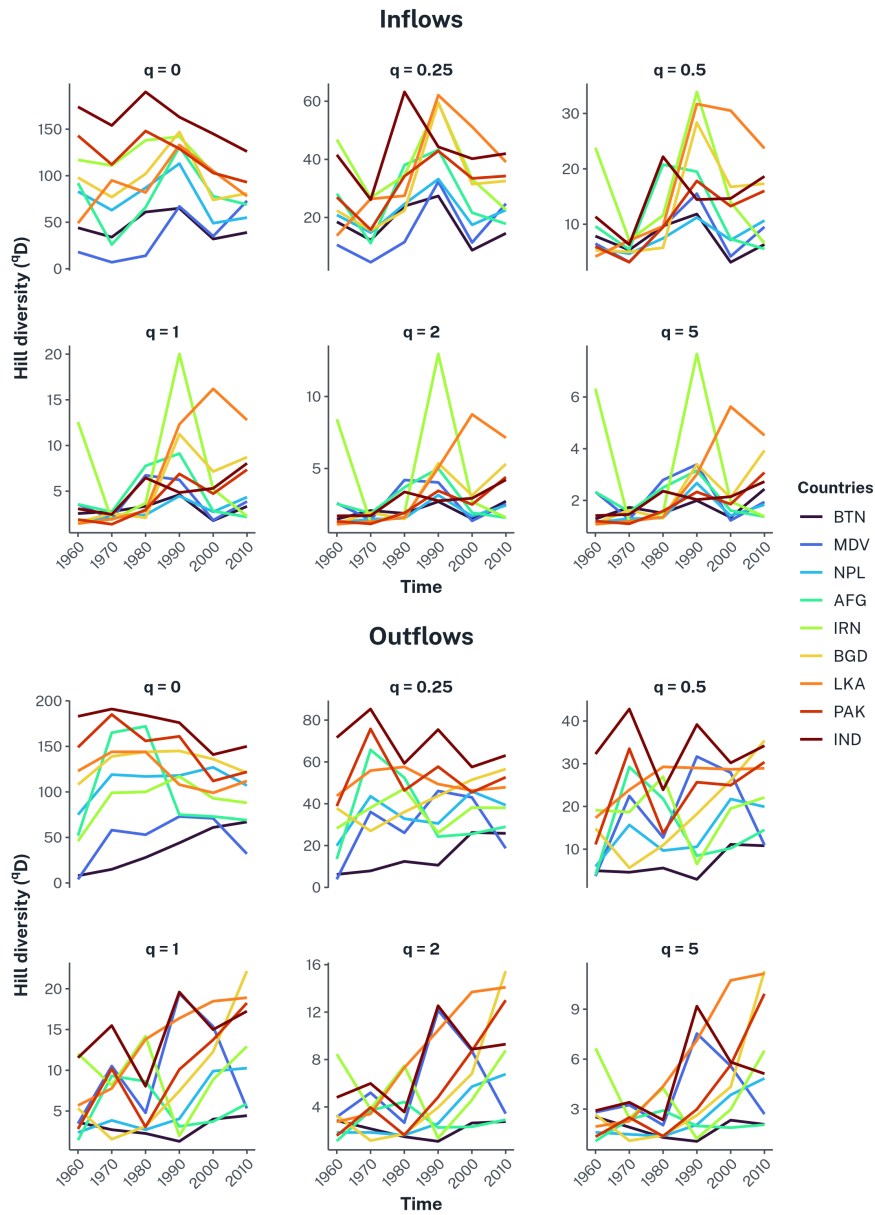


Figure A.6: Migration diversity in inflows and outflows for South Asian countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in Central Asia & Western Asia, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

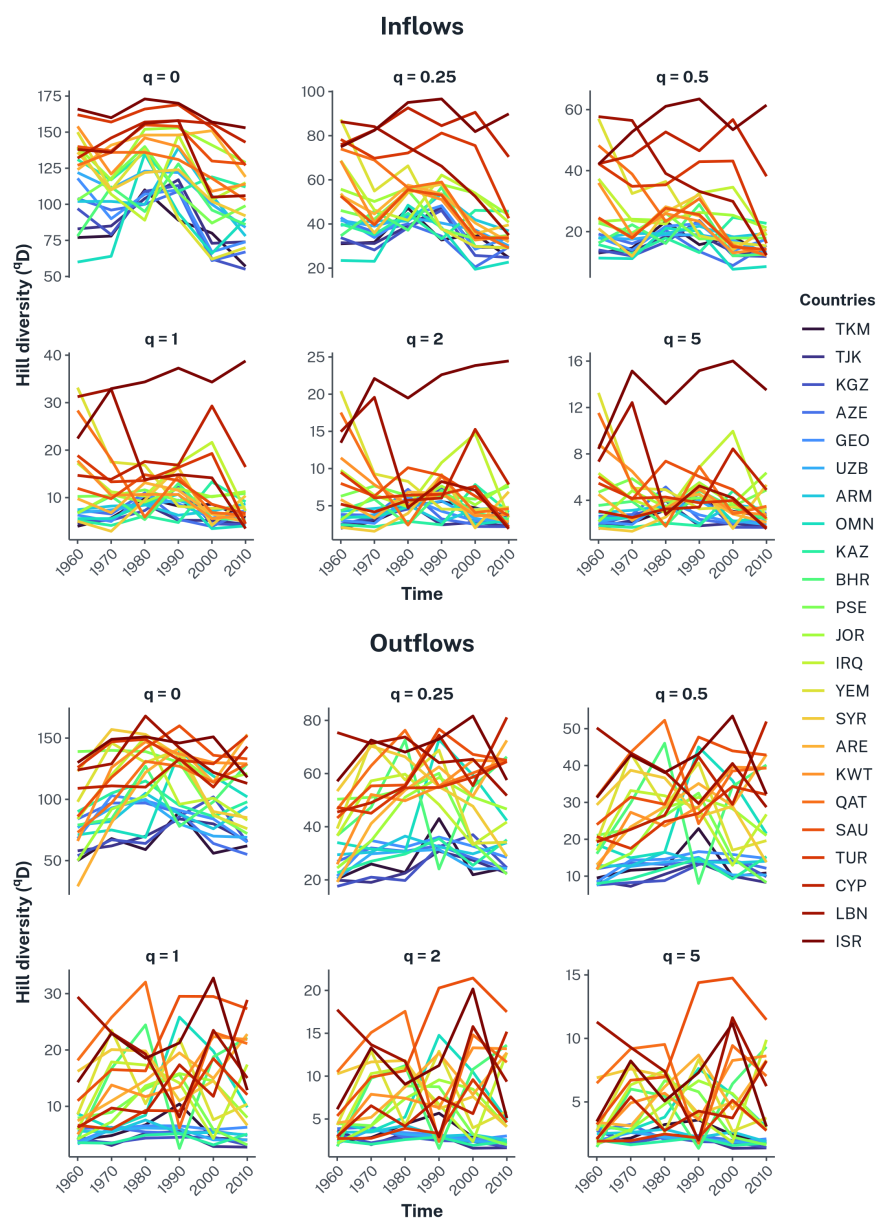


Figure A.7: Migration diversity in inflows and outflows for Central and West Asian countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in Eastern Europe, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

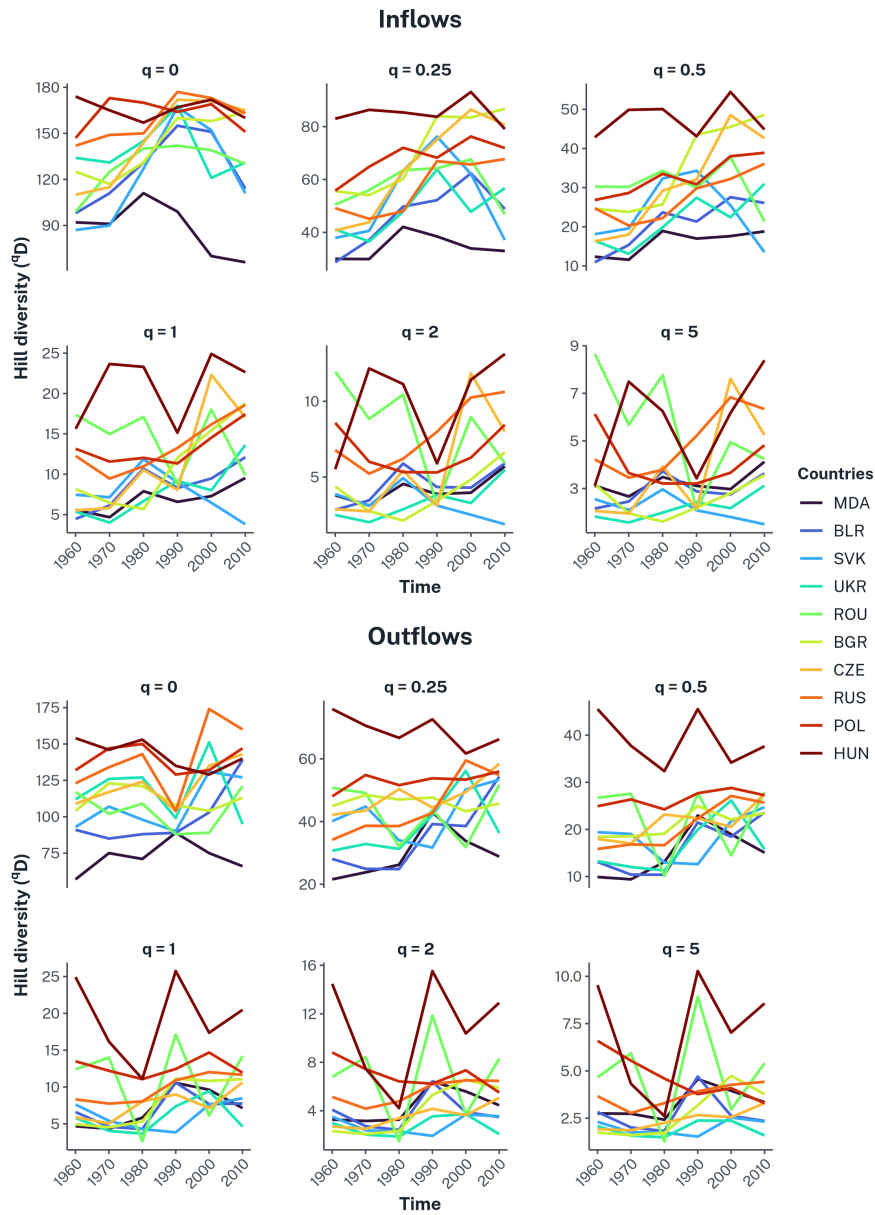


Figure A.8: Migration diversity in inflows and outflows for East European countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in Northern Europe, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

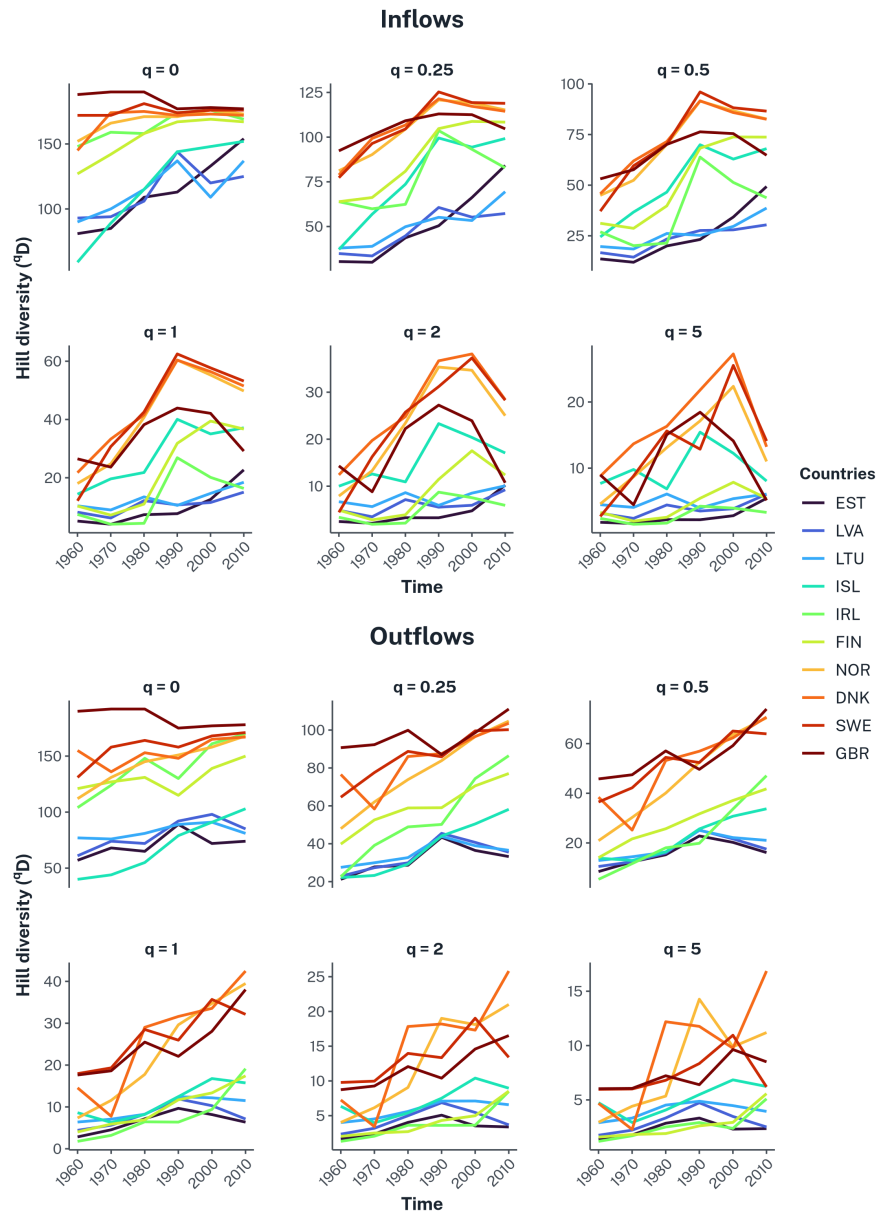


Figure A.9: Migration diversity in inflows and outflows for North European countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in Western Europe, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

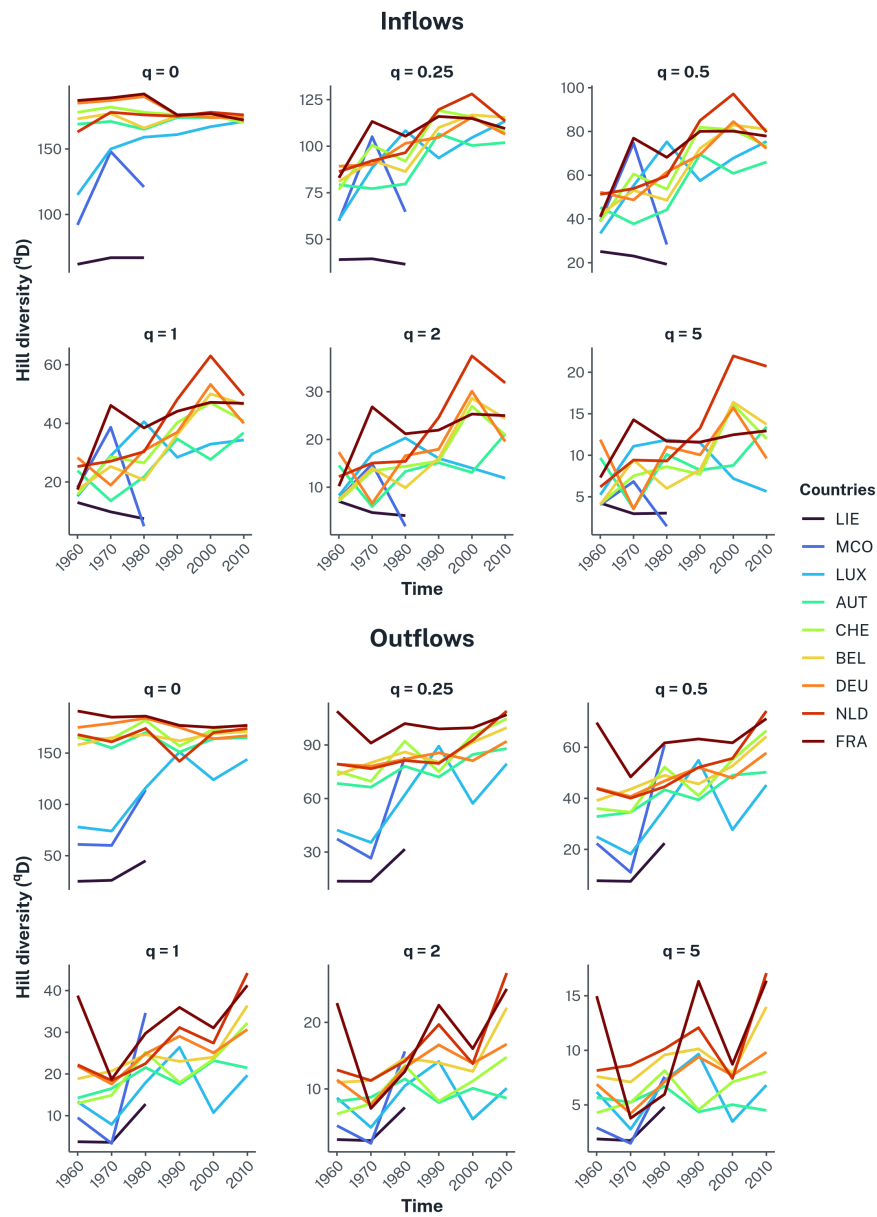


Figure A.10: Migration diversity in inflows and outflows for West European countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in Southern Europe, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

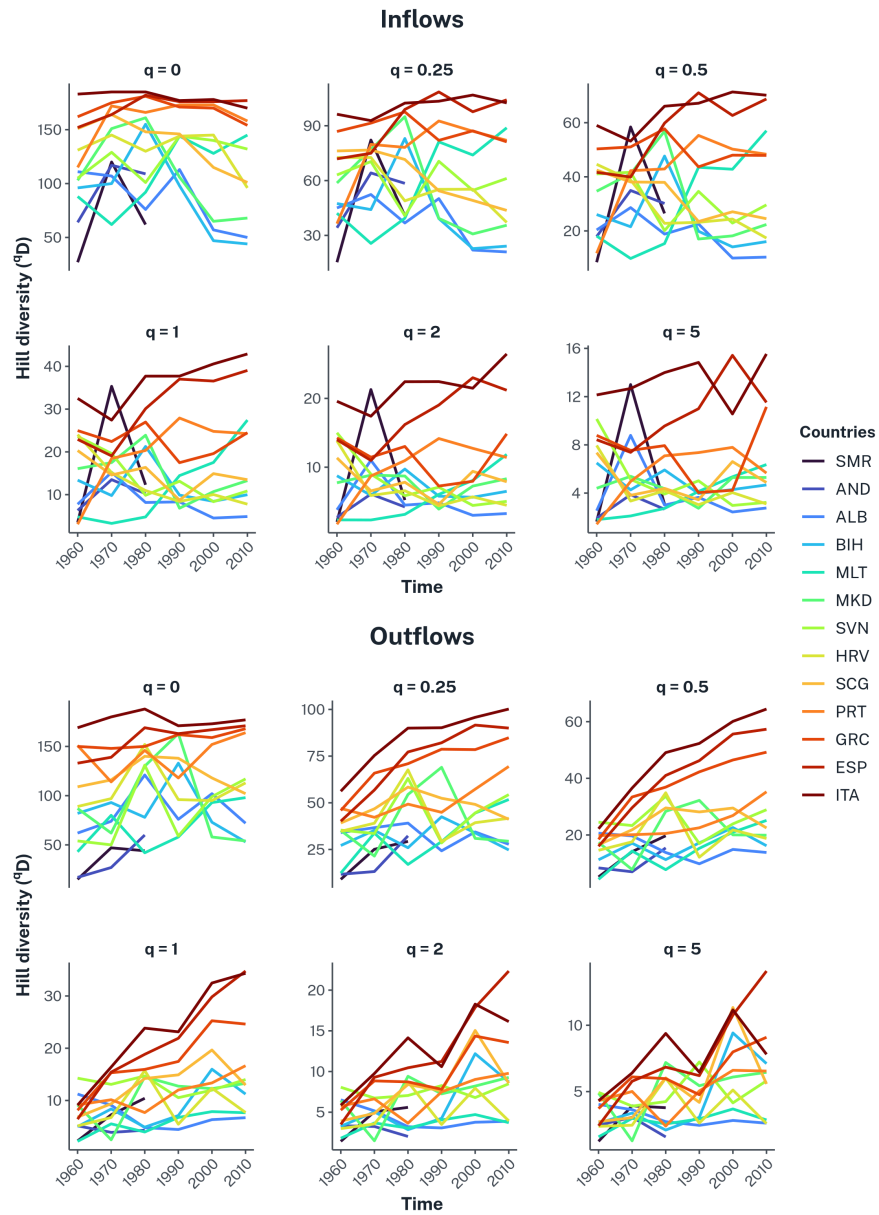


Figure A.11: Migration diversity in inflows and outflows for South European countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in Northern Africa, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

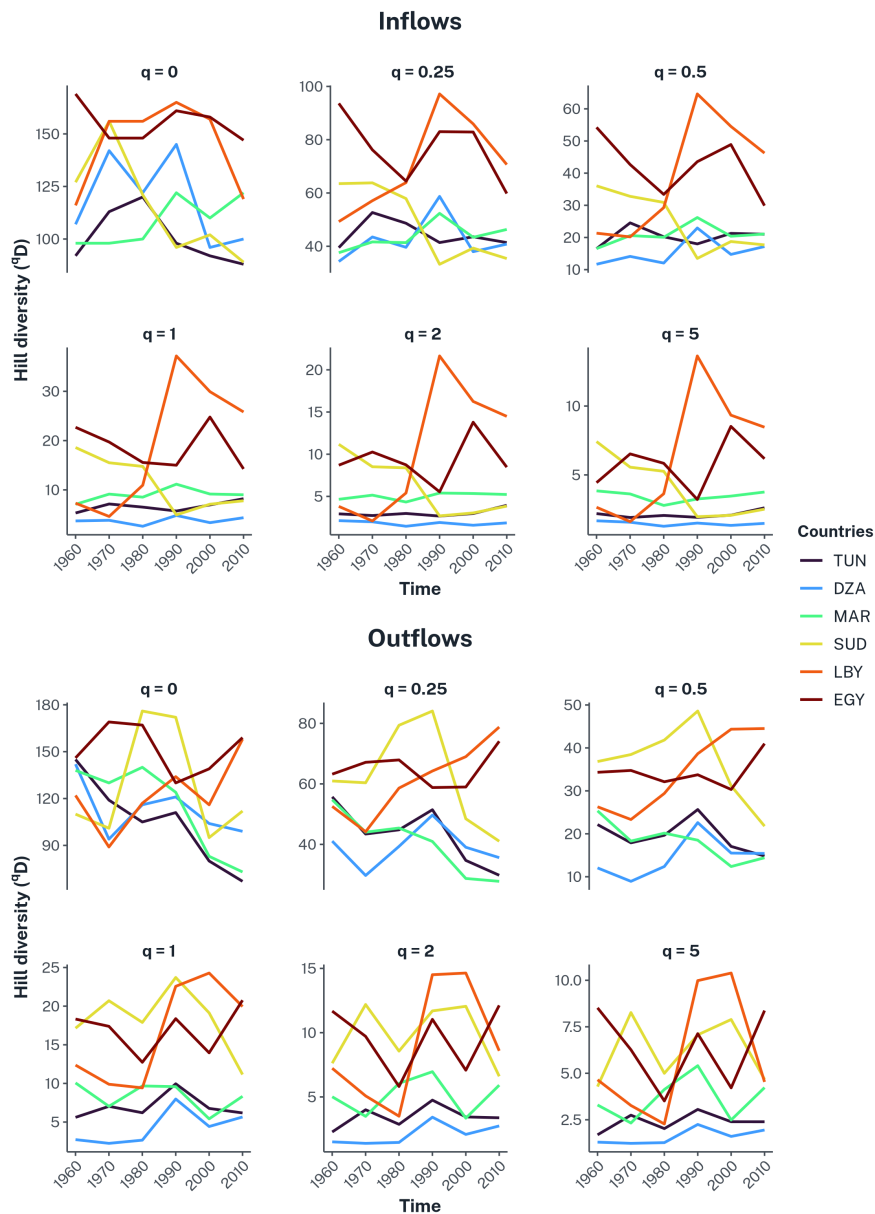


Figure A.12: Migration diversity in inflows and outflows for North African countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in Western Africa, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

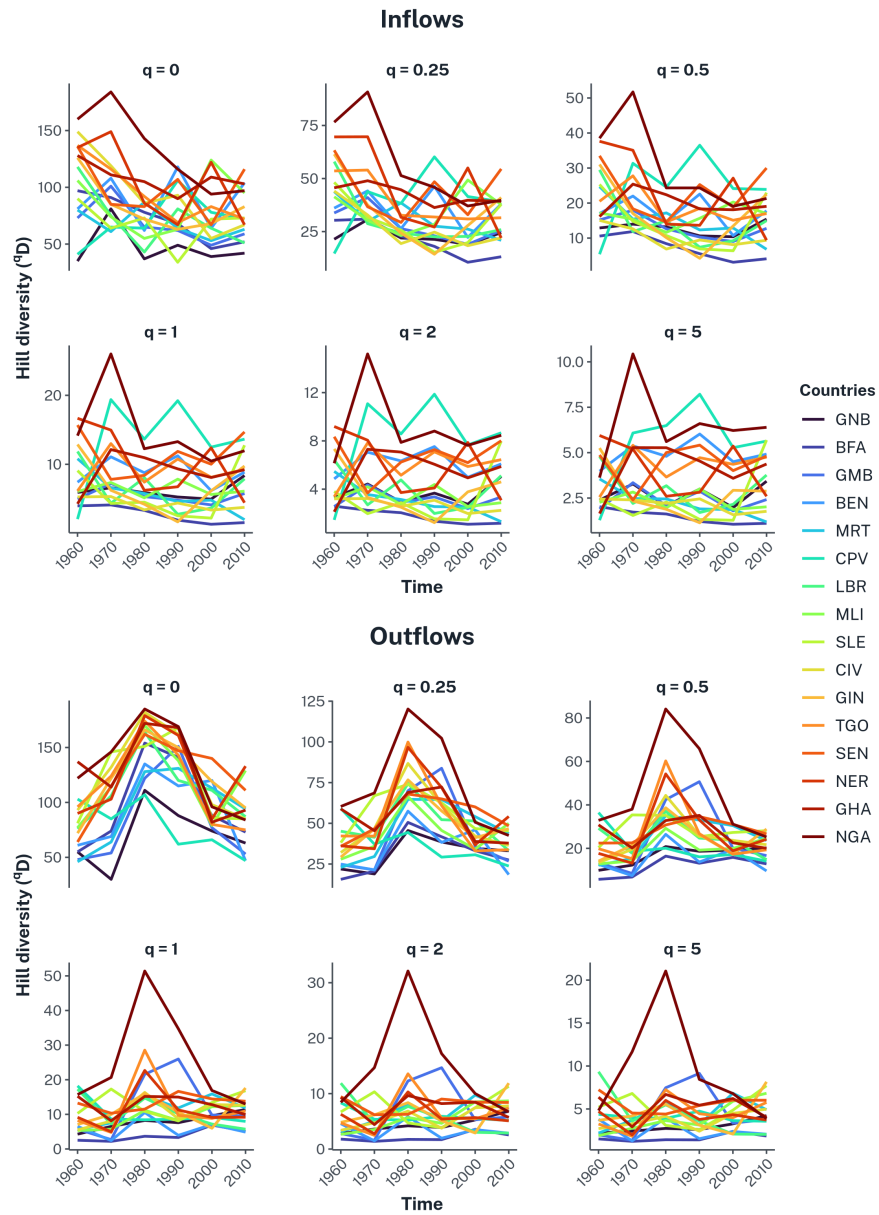


Figure A.13: Migration diversity in inflows and outflows for West African countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in Eastern Africa, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

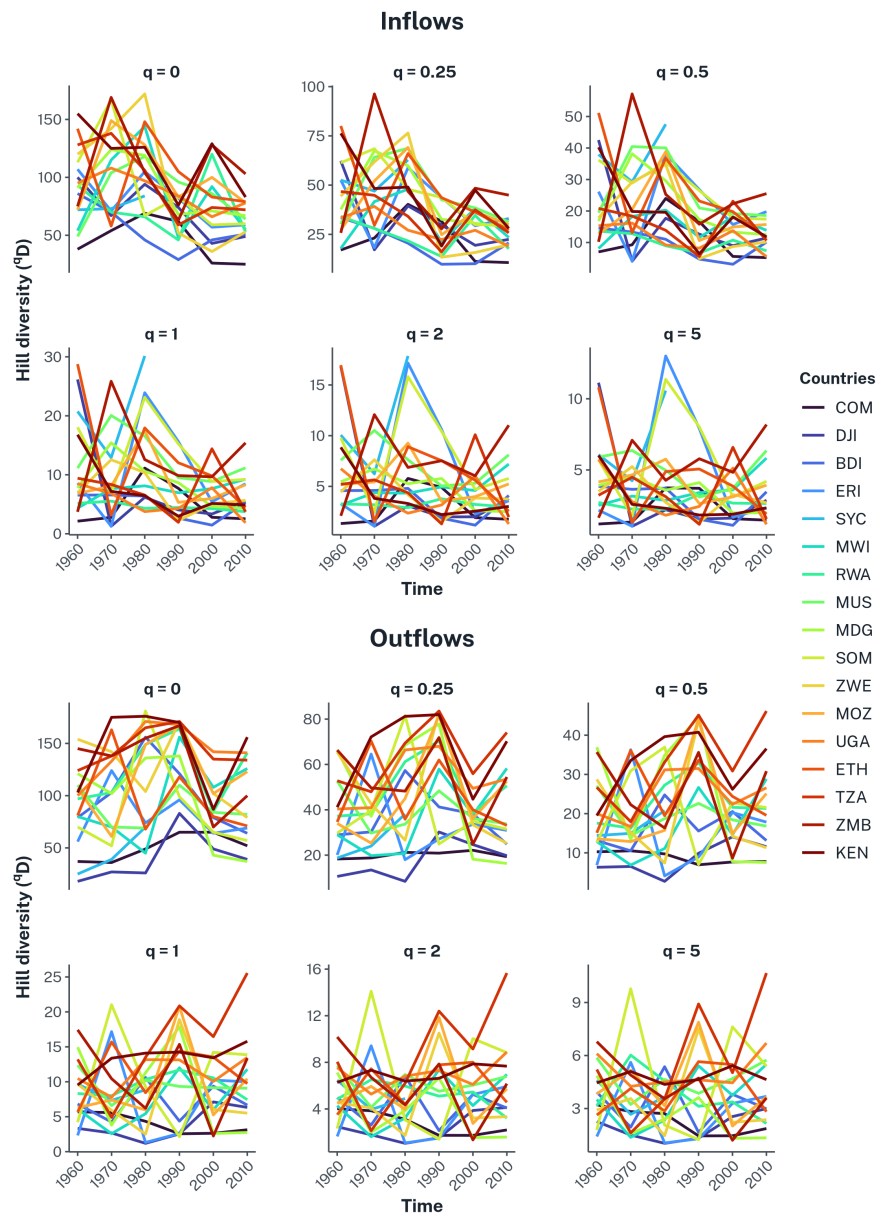


Figure A.14: Migration diversity in inflows and outflows for East African countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in Middle & Southern Africa, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

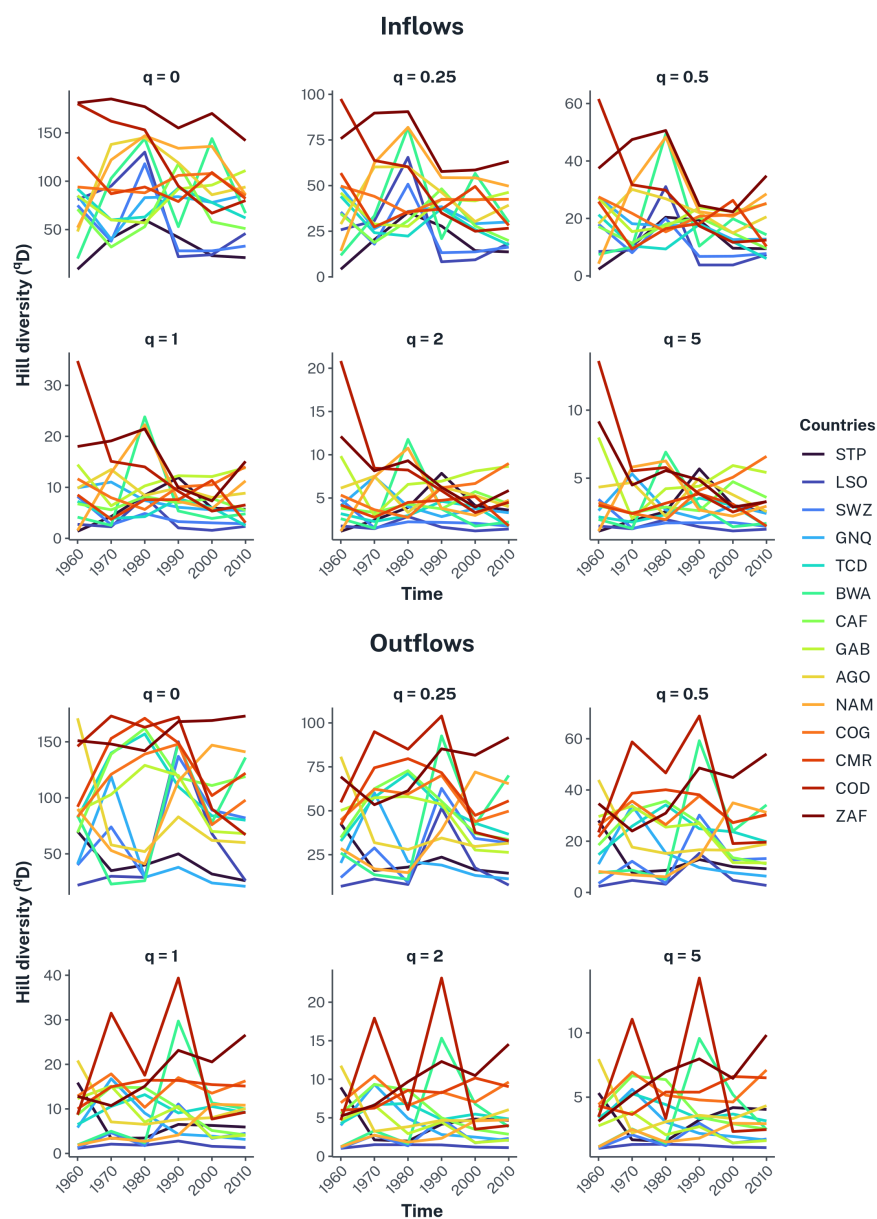


Figure A.15: Migration diversity in inflows and outflows for Middle and South African countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in North America, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

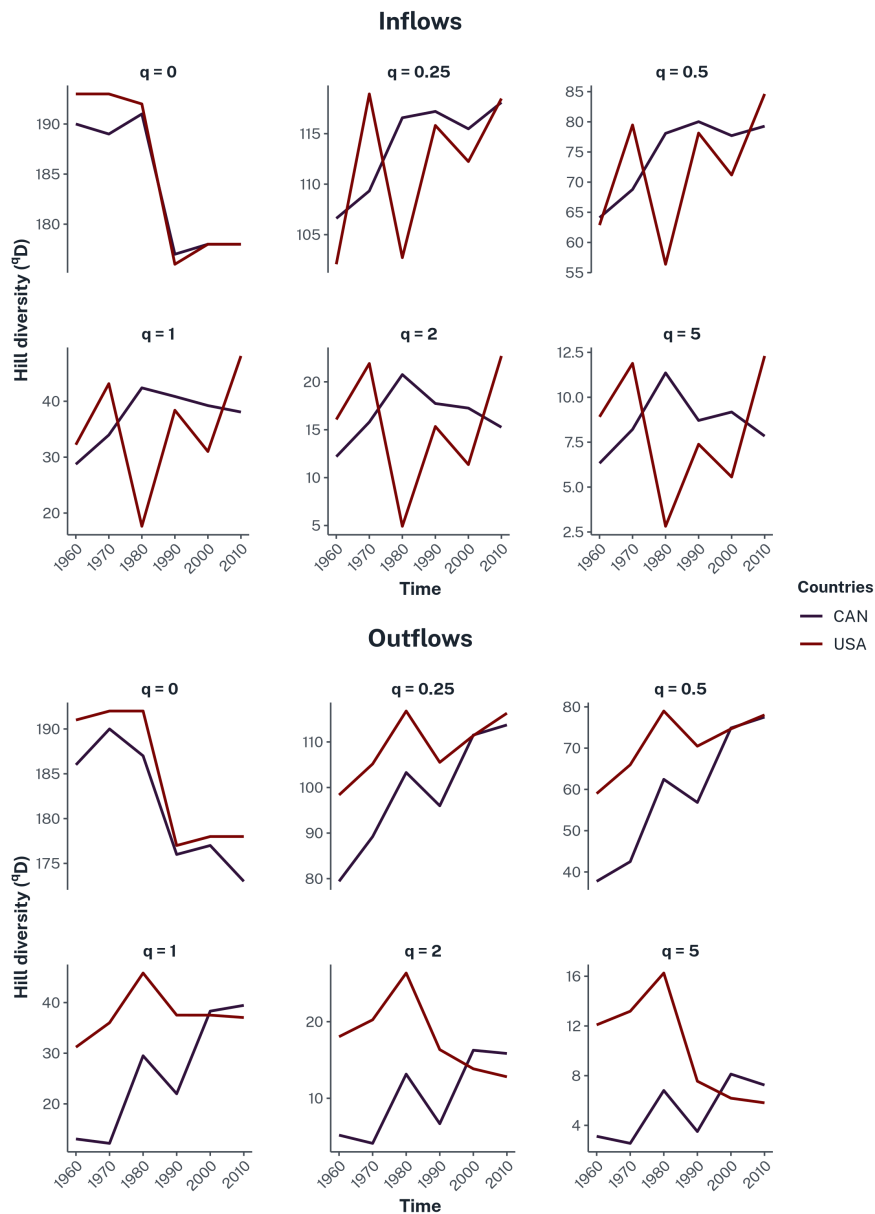


Figure A.16: Migration diversity in inflows and outflows for North American countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in Central America & Caribbean, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

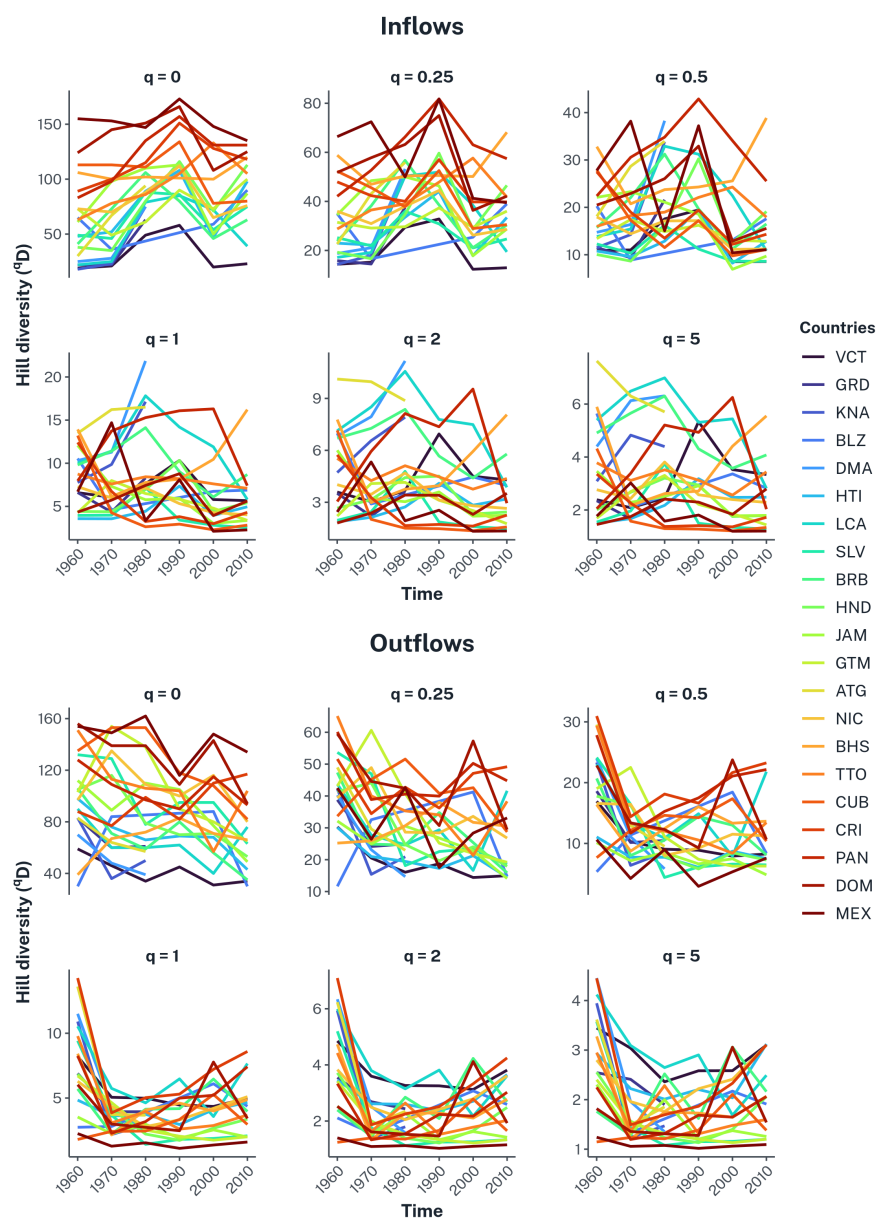


Figure A.17: Migration diversity in inflows and outflows for Central America and Caribbean countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in Southern America, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

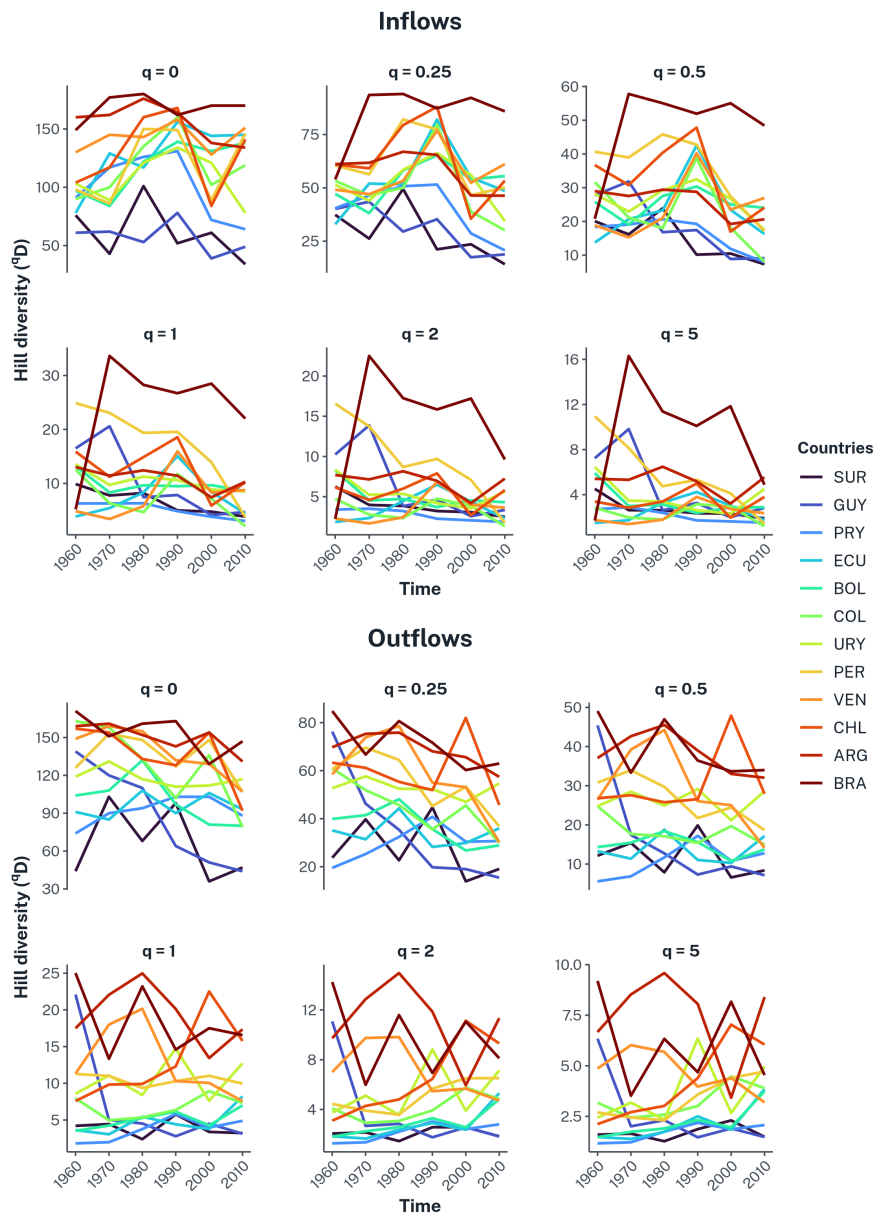


Figure A.18: Migration diversity in inflows and outflows for South American countries, 1960–2020. Each line represents the diversity trajectory of one country.

Changing migration diversity in Oceania, 1960–2020

Countries ranked by average qD across all years, q values, and flow types, from least to most diverse.

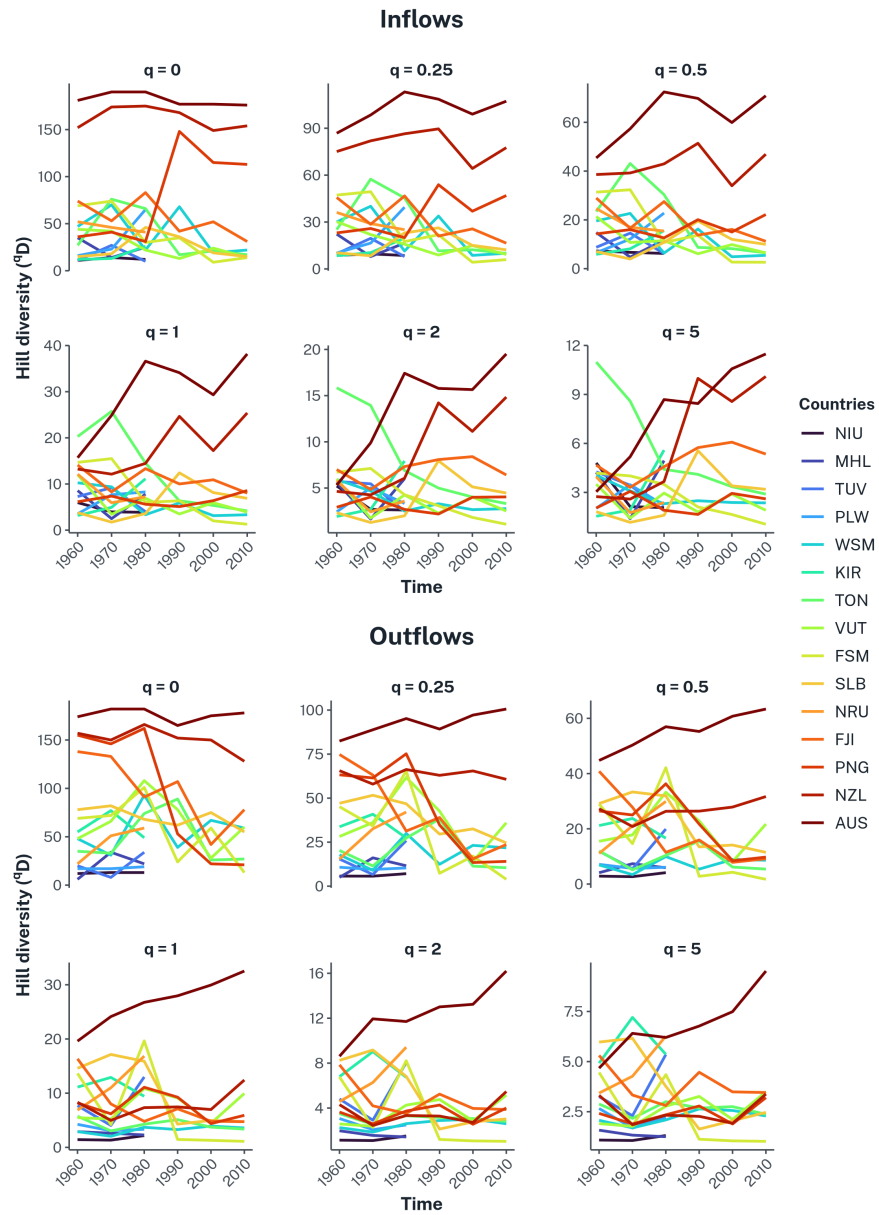


Figure A.19: Migration diversity in inflows and outflows for countries in Oceania, 1960–2020. Each line represents the diversity trajectory of one country.

A.7 Country- and Region-Level Contributions to Connectivity and Evenness over Time

To identify which countries drive global Hill diversity patterns, and whether they operate through connectivity or evenness, I apply a *leave-one-out* strategy. For each country i , I recompute global Hill diversity excluding all corridors in which country i appears as either an origin or a destination, renormalising the remaining flow shares to sum to one. The leave-one-out diversity for country i at order q is:

$${}^qD_{(-i)} = \left(\sum_{j \notin C_i} \tilde{p}_j^q \right)^{1/(1-q)}, \quad (\text{A.1})$$

where C_i denotes the set of corridors involving country i and $\tilde{p}_j = p_j / \sum_{k \notin C_i} p_k$ are the renormalised shares. The contribution of country i to global diversity then is:

$$\delta_i^q = {}^qD - {}^qD_{(-i)}. \quad (\text{A.2})$$

A positive δ_i^q means that country i increases global diversity at order q ; removing it would reduce diversity. A negative value means that country i reduces global diversity, typically because it anchors one or more dominant corridors that concentrate flows and lower evenness. Critically, the sign of δ_i^q can differ across q : a country may contribute positively to connectivity while contributing negatively to evenness, if it maintains many active corridors including rare ones but also anchors a dominant corridor that suppresses higher-order diversity. This profile of contributions across q makes it possible to characterise each country's structural role in the global migration system and to identify which countries are driving corridor turnover, flow redistribution, or both. Country-level contributions are aggregated to the regional level and tracked across all six periods to examine how the geographic drivers of structural change have shifted over time.

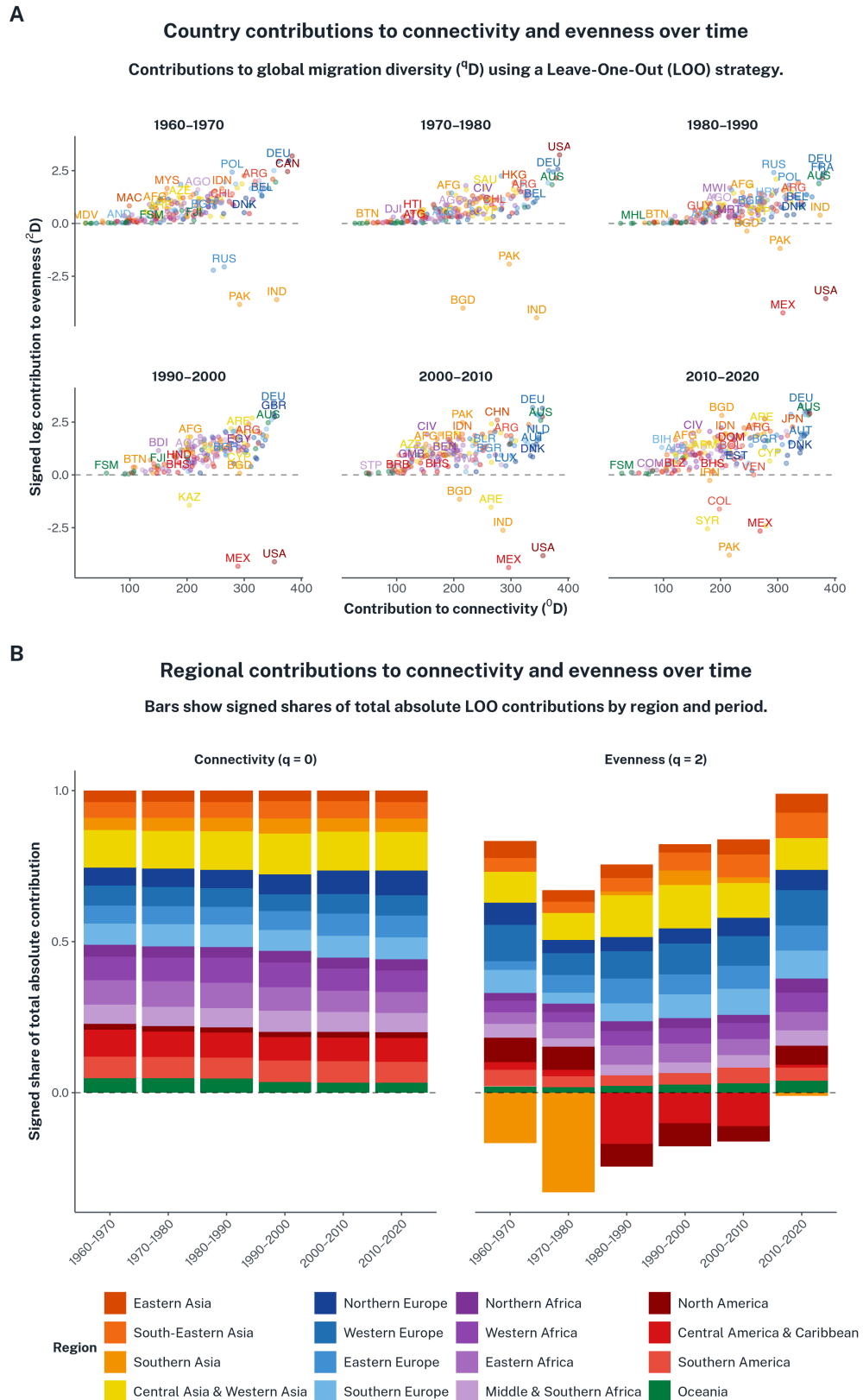


Figure A.20: Country- and region-level contributions to connectivity and evenness over time, using pseudo-Bayesian flow estimates.

A.8 Replication with Minimisation–Closed Flow Estimation

To ensure the robustness of the findings, I replicate the empirical analysis using the minimisation–closed demographic accounting method to estimate all country-to-country migration flows.

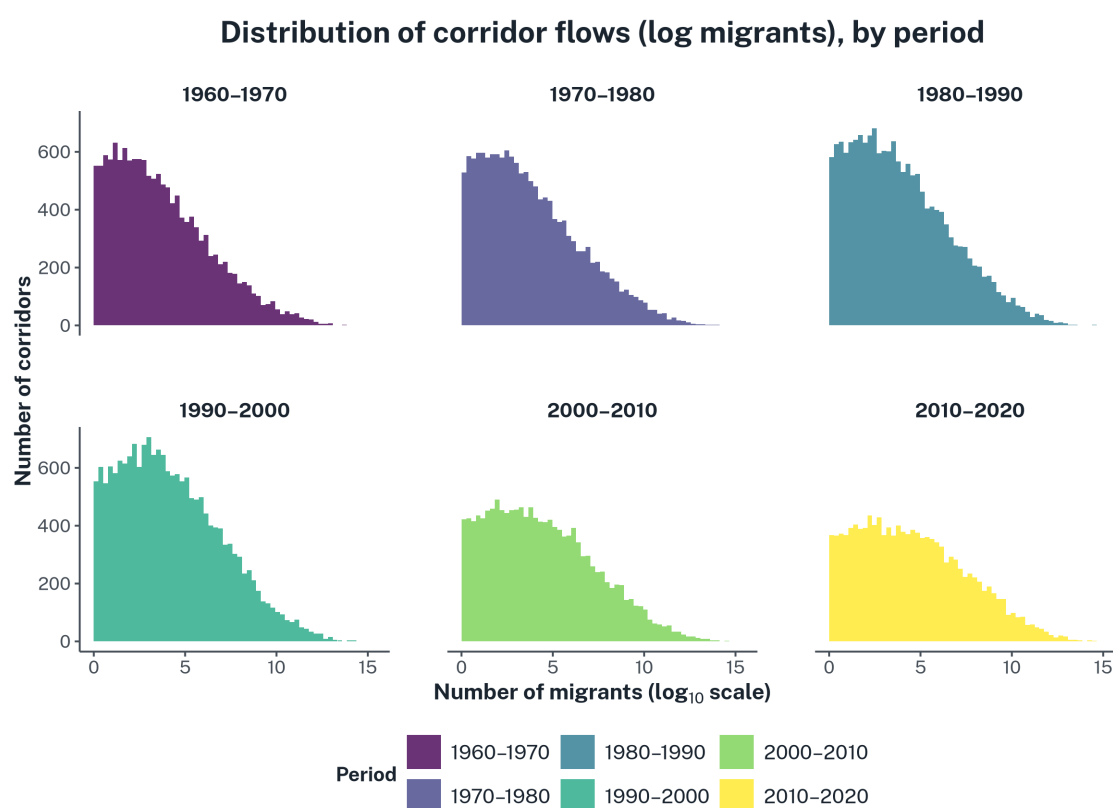


Figure A.21: Distribution of corridor volumes over time, 1960–2020, using minimisation–closed flow estimates.

Country-level migration inflows and outflows

Total inflows and outflows by country, using minimisation (closed population) flow estimation. Robinson projection centred on 0° longitude.

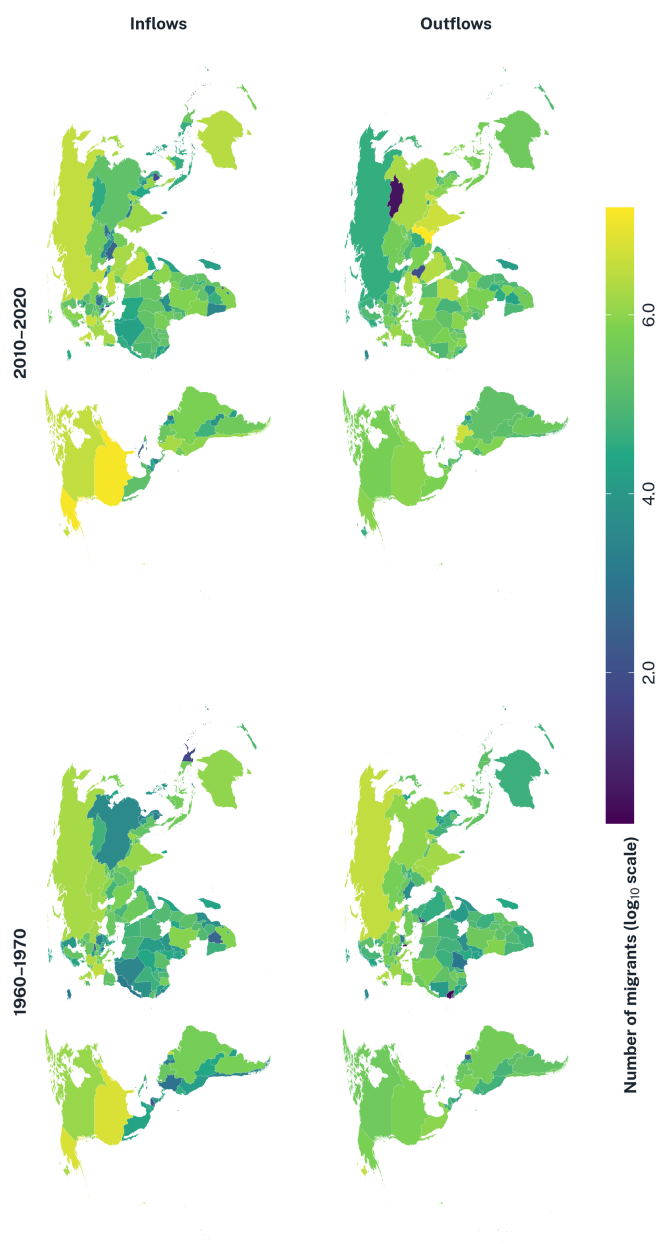


Figure A.22: Choropleth maps of country-level migration inflows and outflows in the first (1960–1970) and last (2010–2020) periods under study, using minimisation–closed flow estimates. Values are shown on a $\log_{10}$ scale. Maps use the Robinson projection centred on 0° longitude.

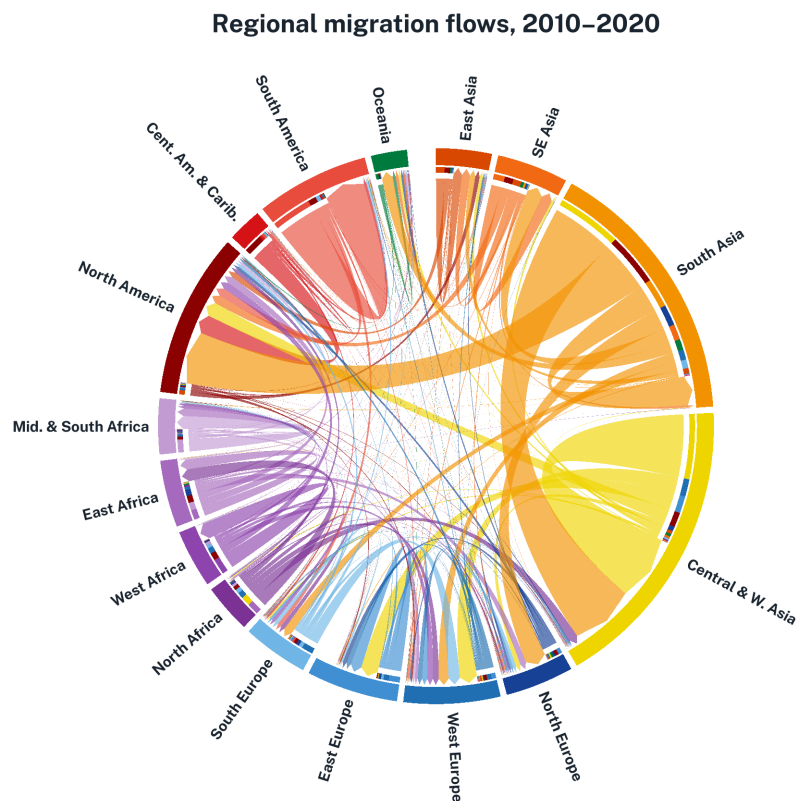
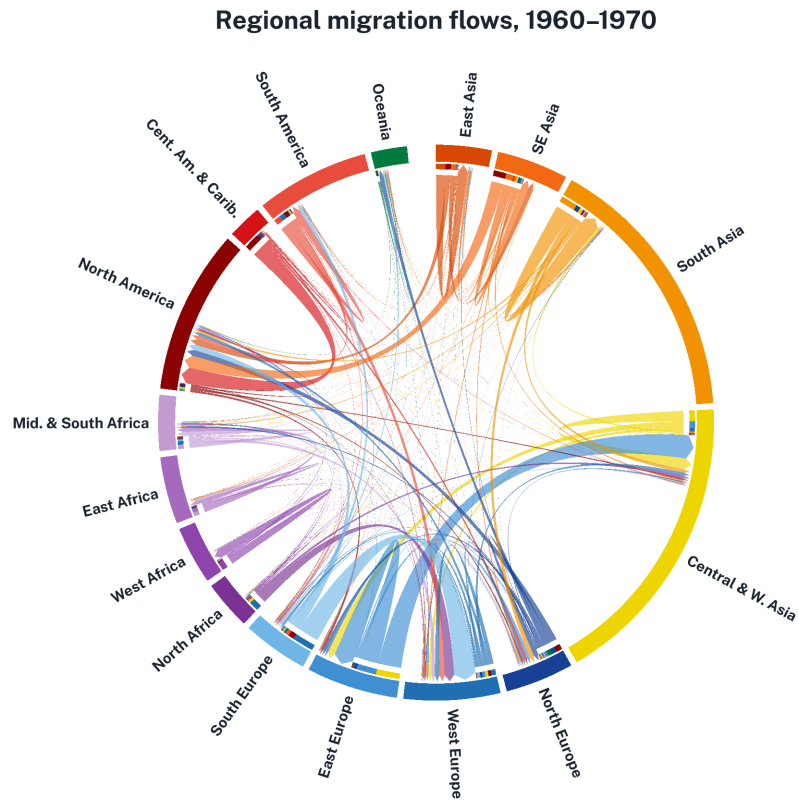


Figure A.23: Chord diagrams of region-to-region migration flows in the first (1960–1970) and last (2010–2020) periods under study, using minimisation–closed flow estimates.

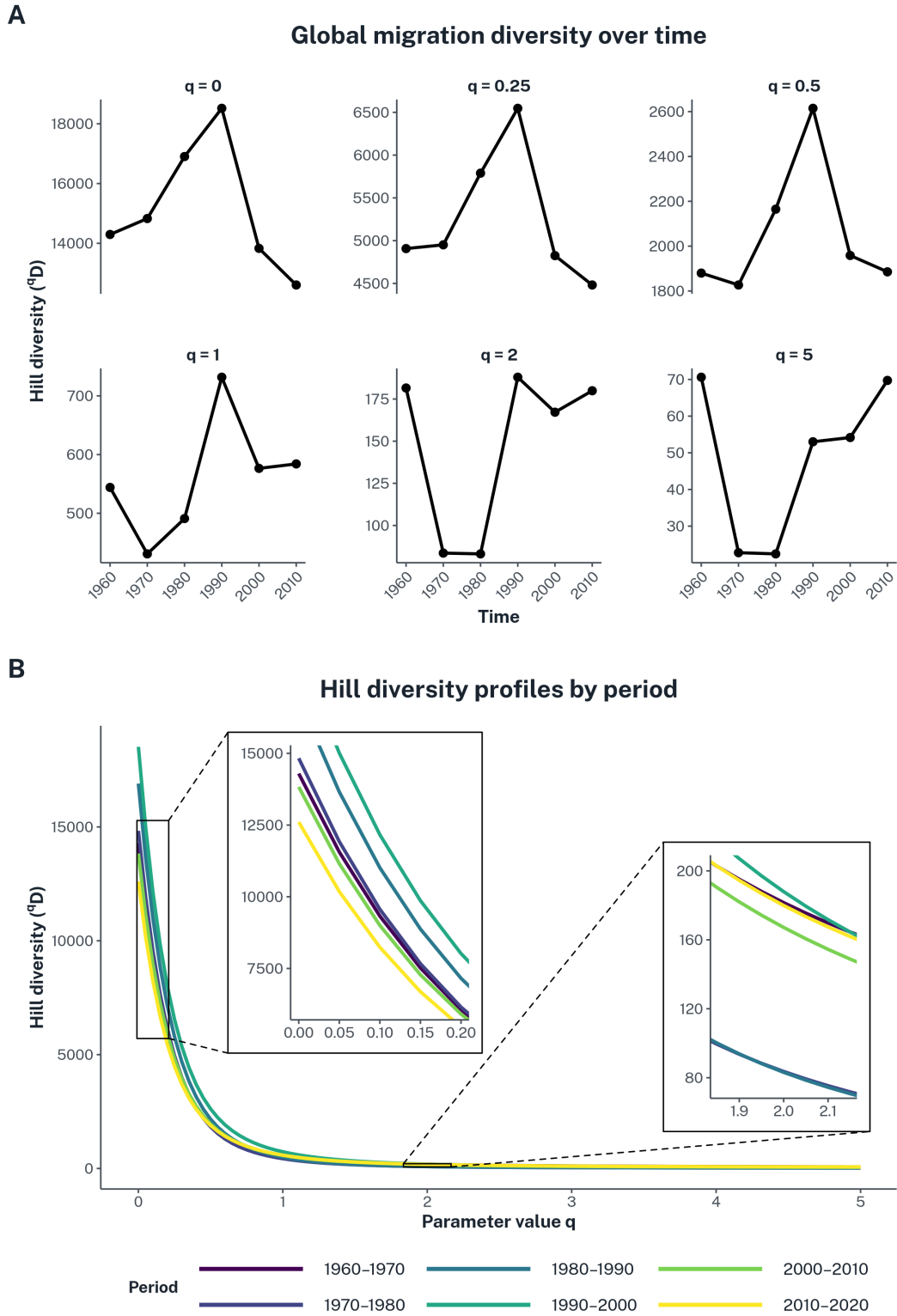
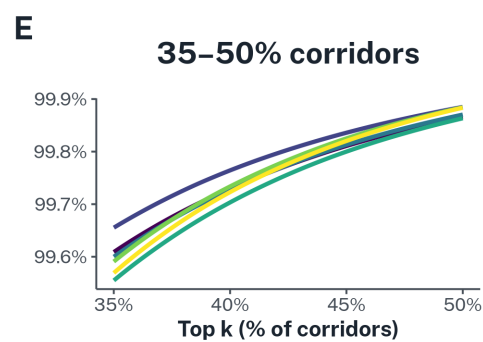
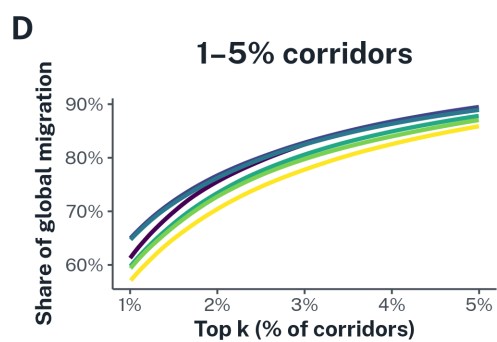
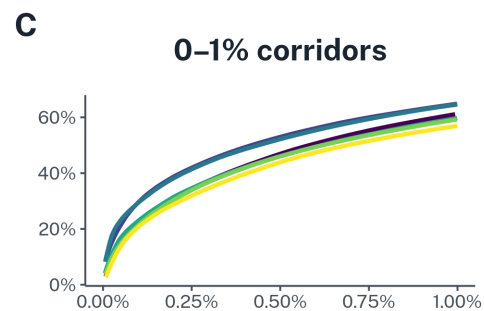
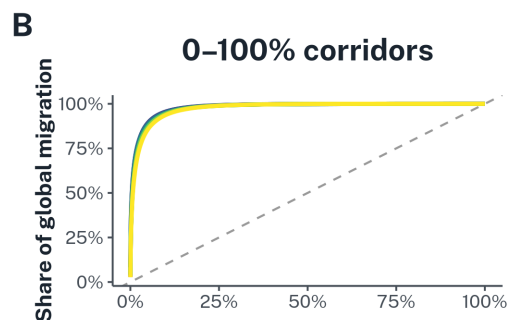
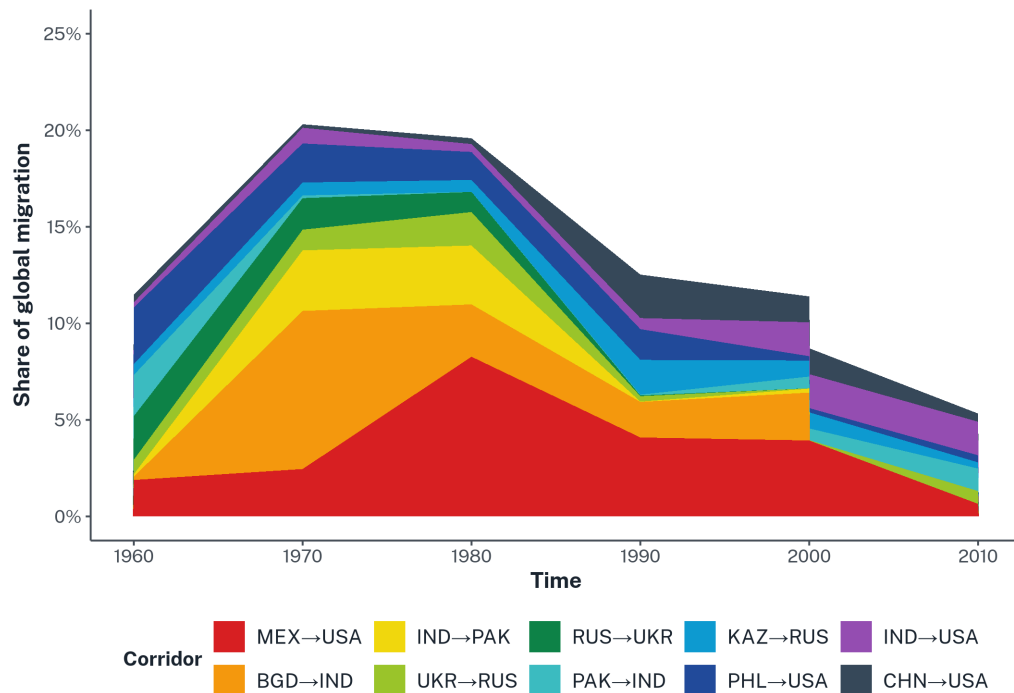


Figure A.24: Hill diversity of the global migration system, 1960–2020, using minimisation–closed flow estimates. Panel A reports the Hill numbers for selected q values over time. Panel B reports the entire Hill profiles of each period, shown as differently-coloured lines.

A Share of global migration in the top 10 corridors (ranked by total flows)



Period

- 1960–1970
- 1970–1980
- 1980–1990
- 1990–2000
- 2000–2010
- 2010–2020

Figure A.25: Evolution of the dominance of the most important corridors in the global migration system, 1960–2020, using minimisation–closed flow estimates.

Changes in global migration diversity under observed vs stable corridor set

Diversity change in observed system vs counterfactual with no corridor turnover.

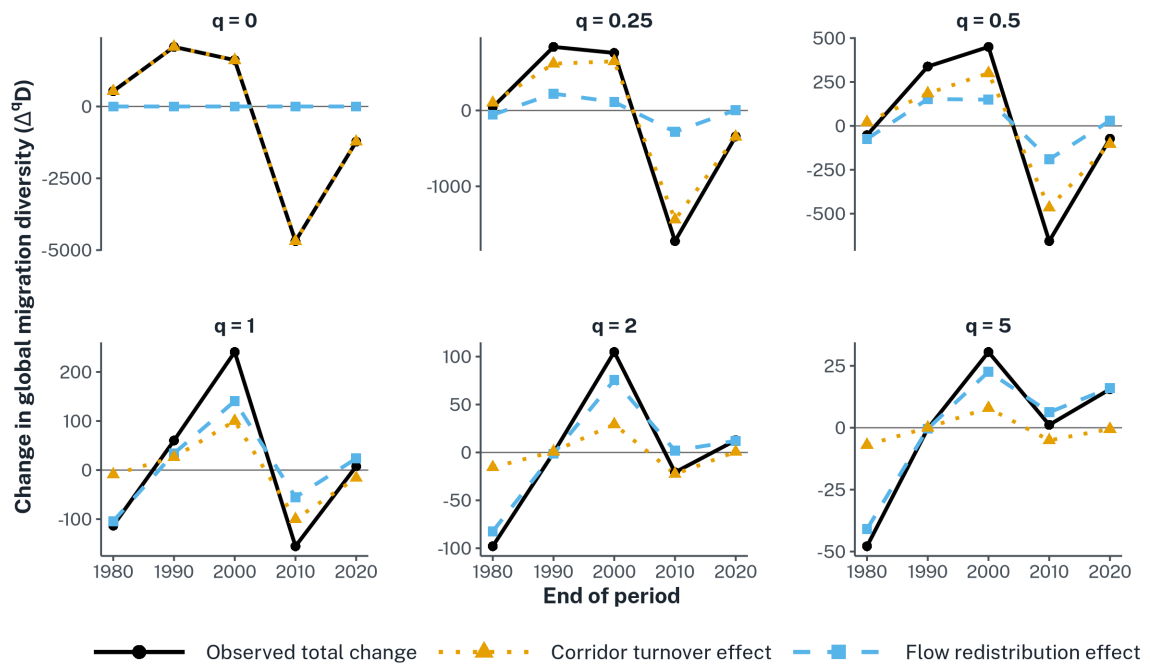


Figure A.26: Comparison of changes in global migration diversity through a stable-support analysis, using minimisation-closed flow estimates.

Within- and between-country sources of migration diversity

Partitioning of system-level Hill diversity (qD) into alpha and beta components.

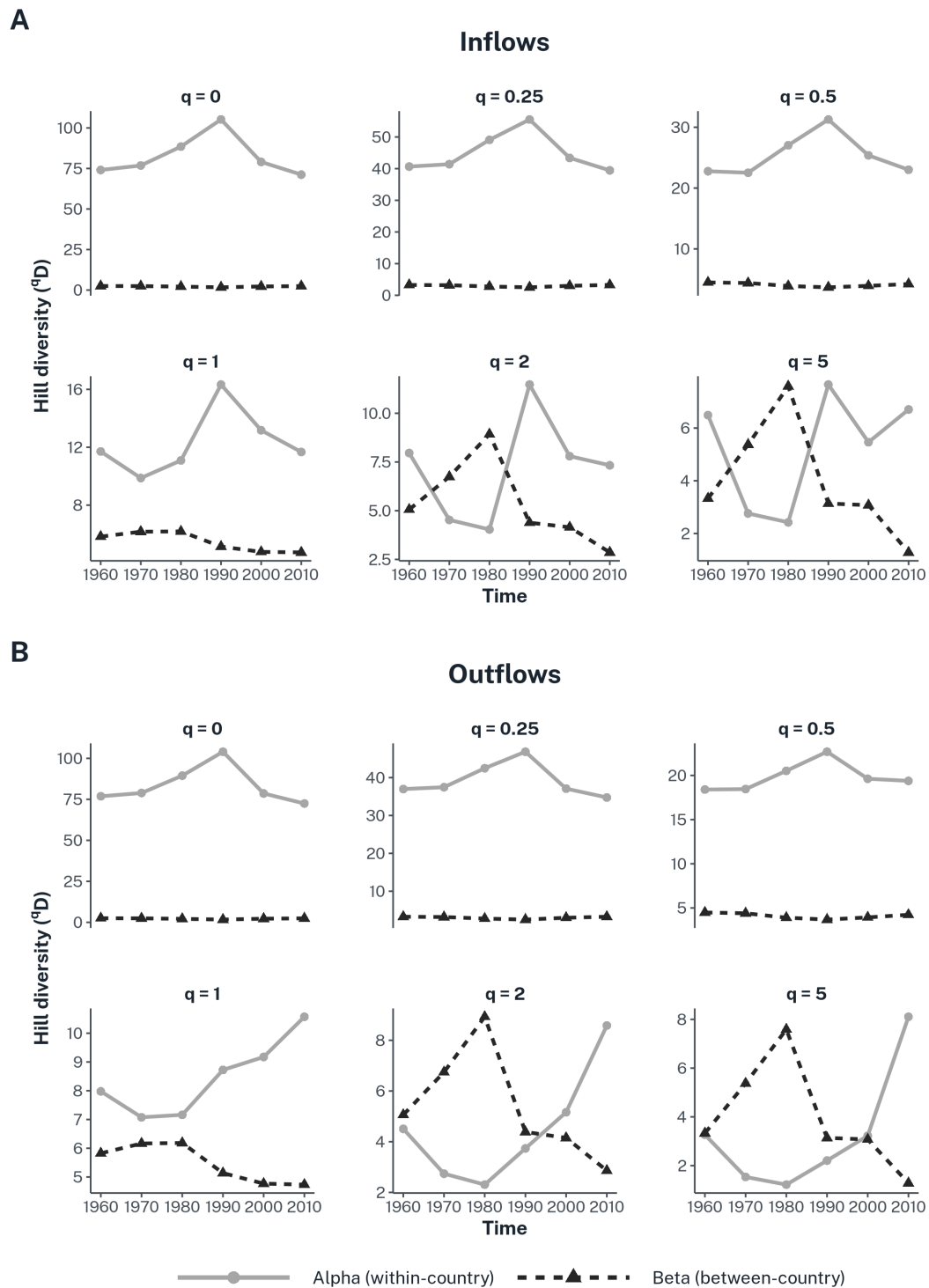
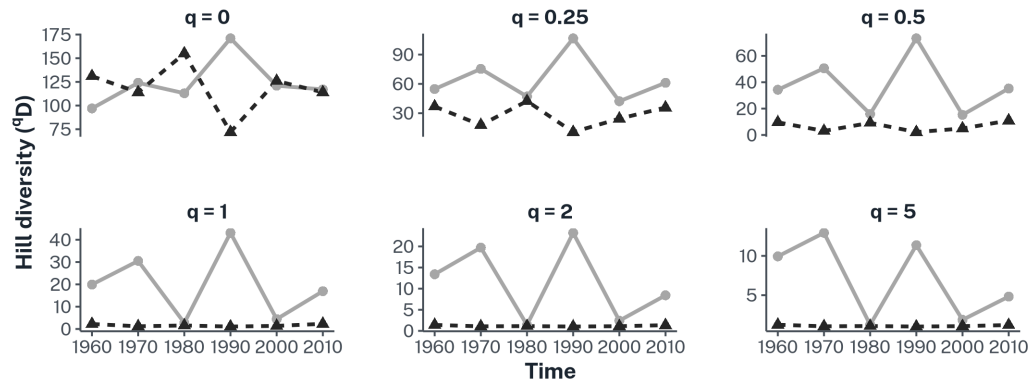


Figure A.27: Partitioning of global migration diversity into within- and between-country components, using minimisation–closed flow estimates.

Diversity of migration inflows and outflows for selected countries

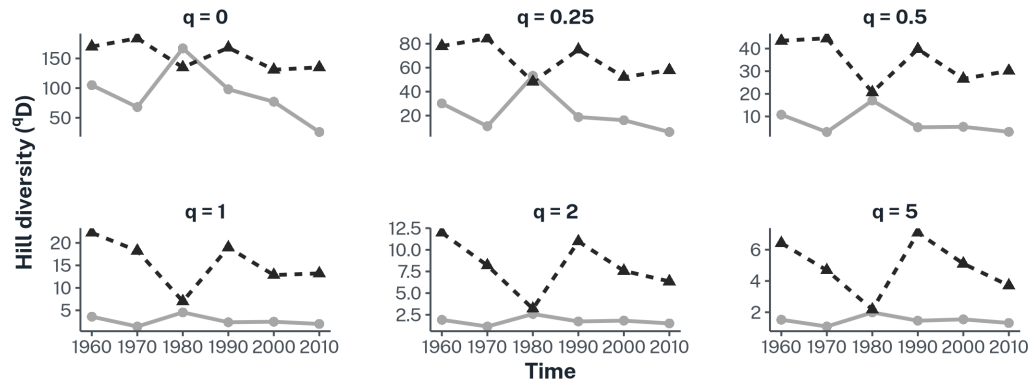
A

Mexico



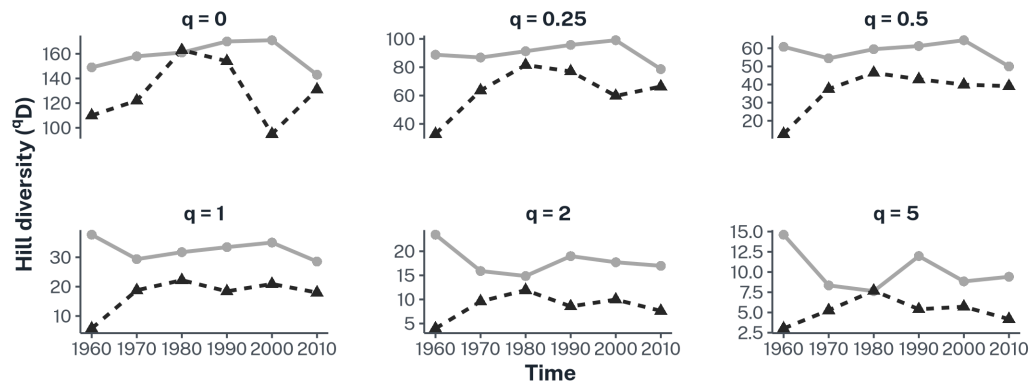
B

India



C

Italy



Flow type —●— Inflows - -▲- - Outflows

Figure A.28: Hill diversity of migration inflows and outflows for selected countries, 1960–2020, using minimisation–closed flow estimates.

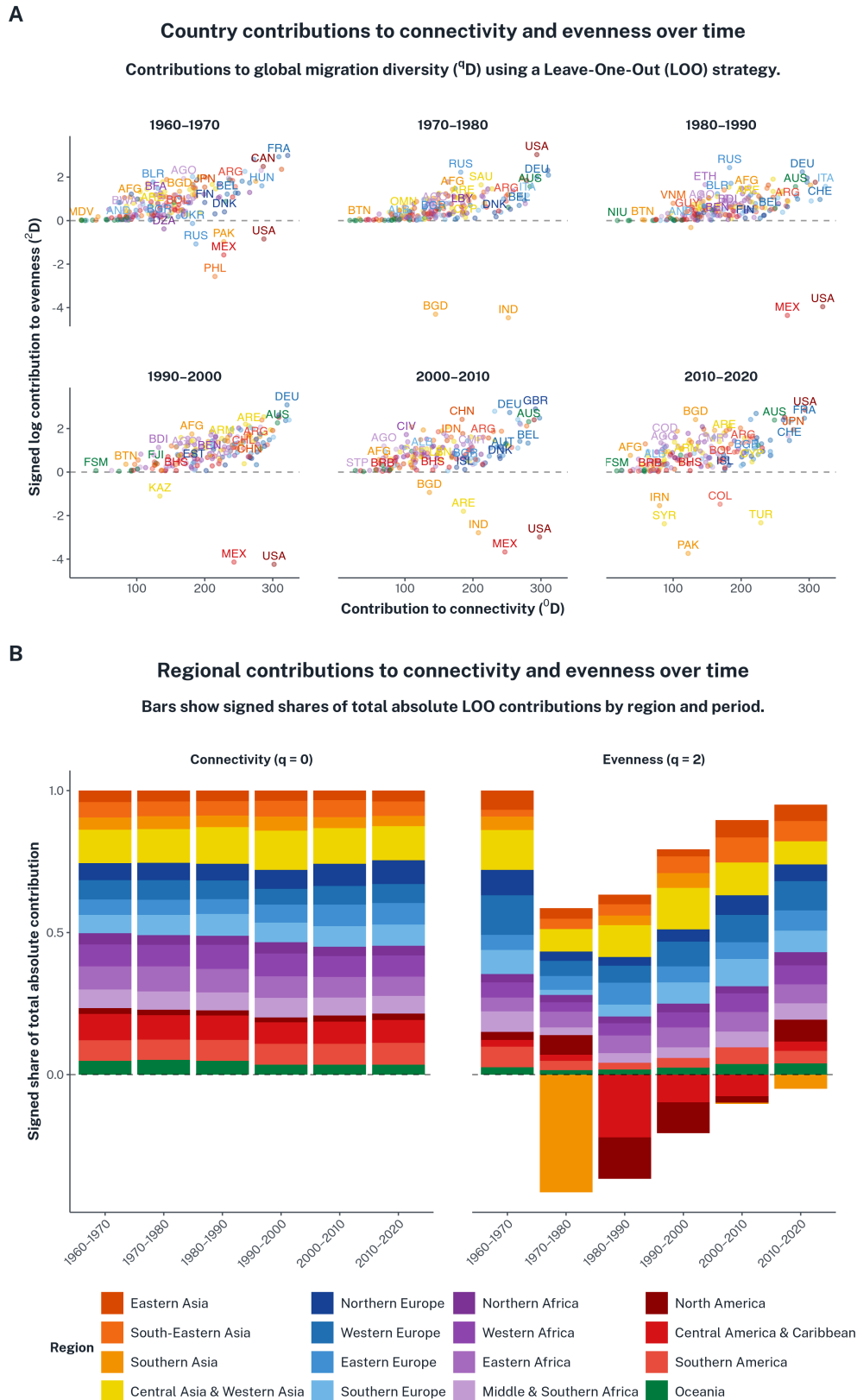


Figure A.29: Country- and region-level contributions to connectivity and evenness over time, using minimisation–closed flow estimates.

A.9 Replication with Minimisation–Open Flow Estimation

To ensure the robustness of the findings, I replicate the empirical analysis using the minimisation–open demographic accounting method to estimate all country-to-country migration flows.

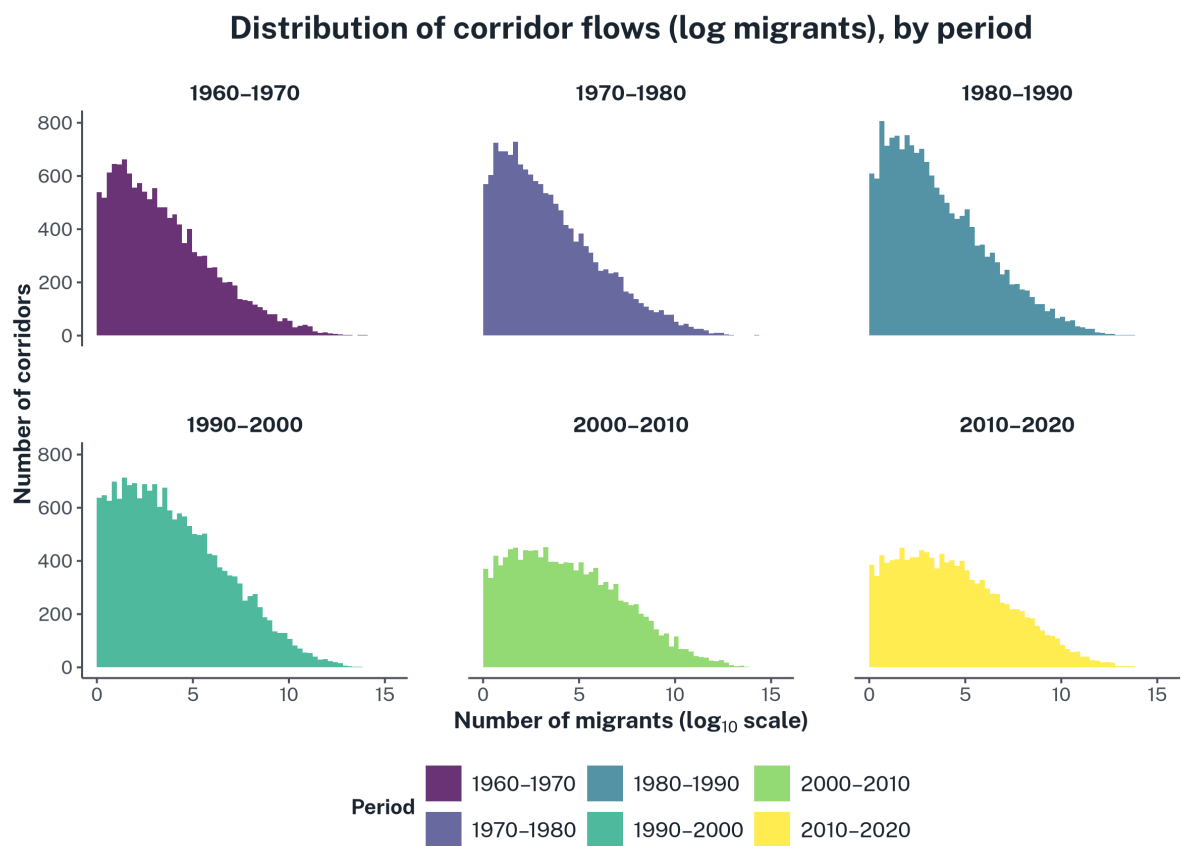


Figure A.30: Distribution of corridor volumes over time, 1960–2020, using minimisation–open flow estimates.

Country-level migration inflows and outflows

Total inflows and outflows by country, using minimisation (open population) flow estimation. Robinson projection centred on 0° longitude.

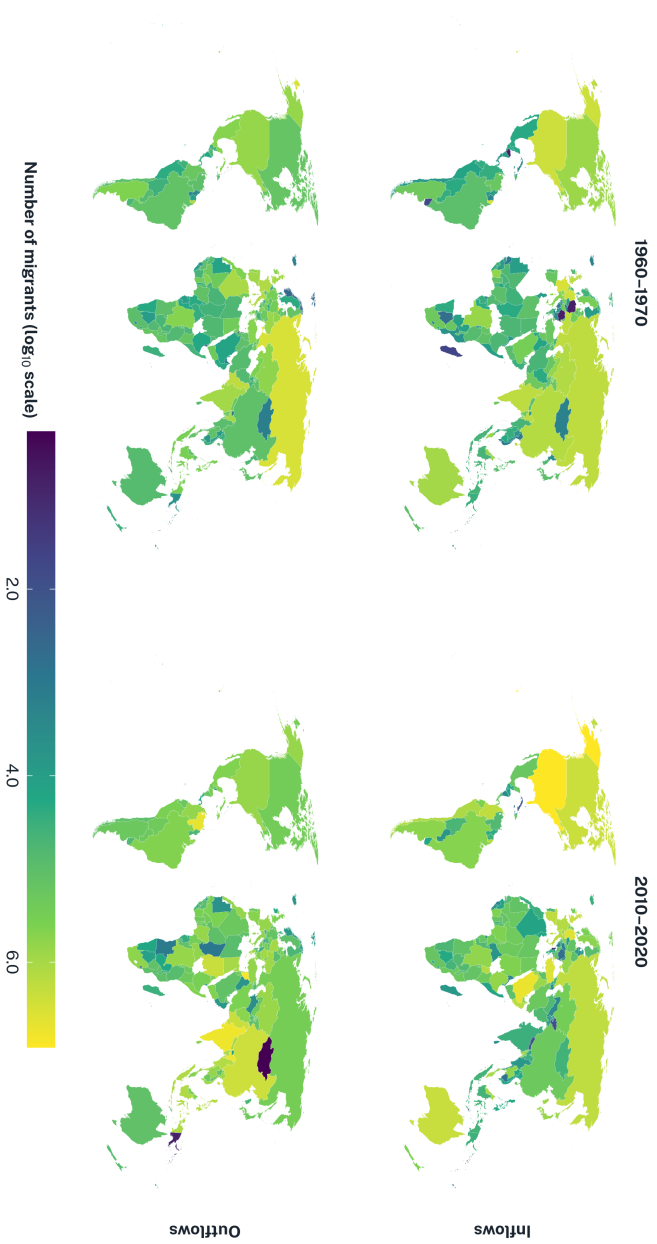


Figure A.31: Choropleth maps of country-level migration inflows and outflows in the first (1960–1970) and last (2010–2020) periods under study, using minimisation–open flow estimates. Values are shown on a log₁₀ scale. Maps use the Robinson projection centred on 0° longitude.

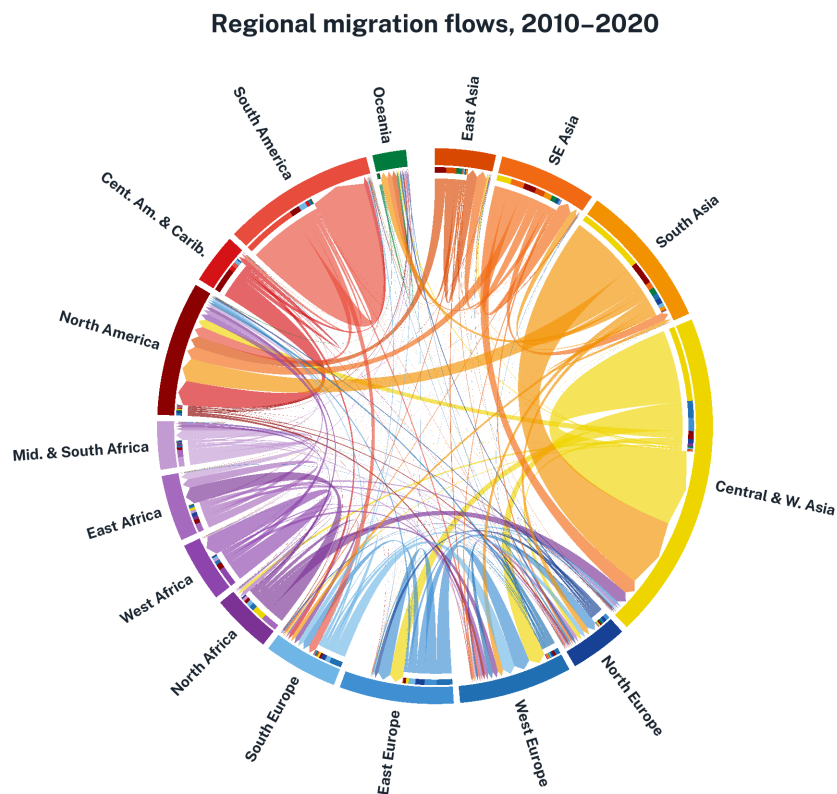
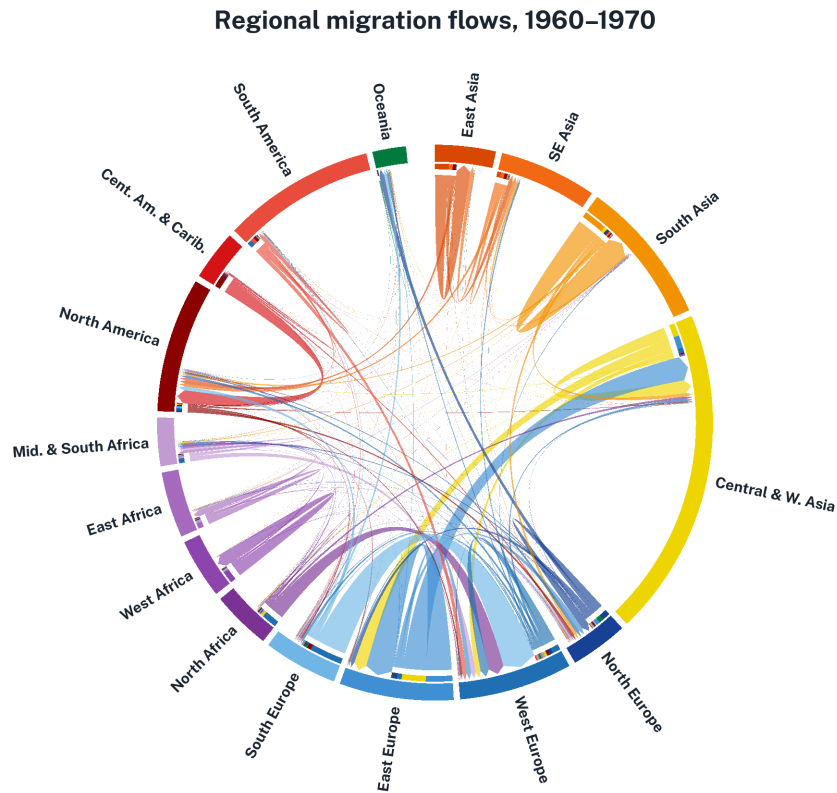


Figure A.32: Chord diagrams of region-to-region migration flows in the first (1960–1970) and last (2010–2020) periods under study, using minimisation–open flow estimates.

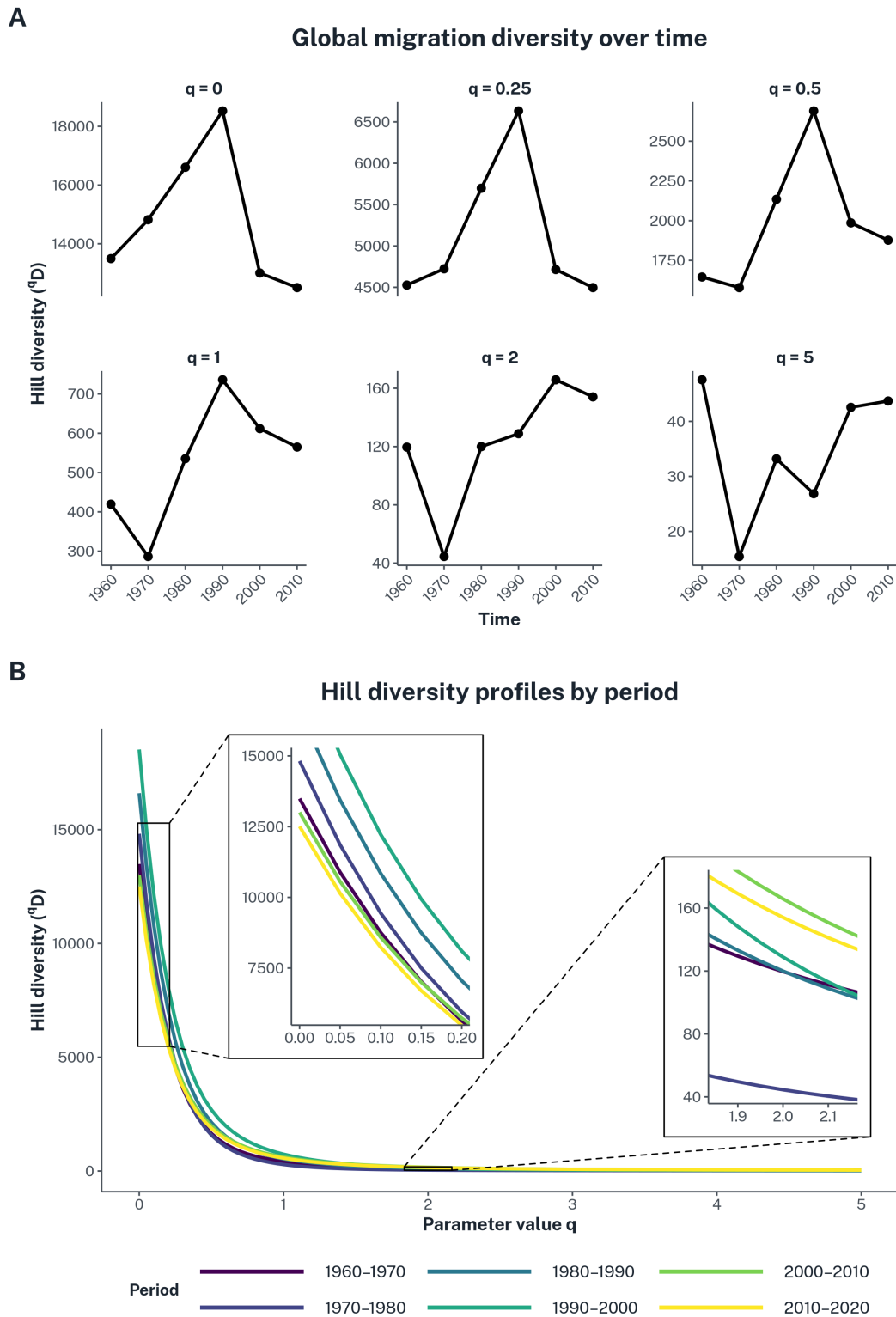
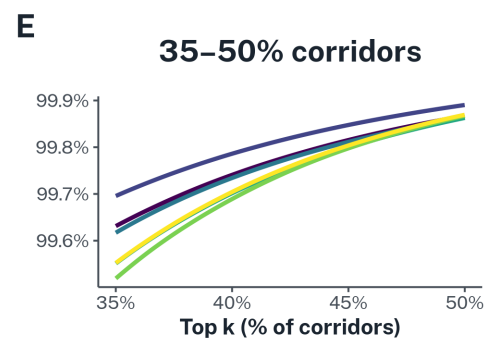
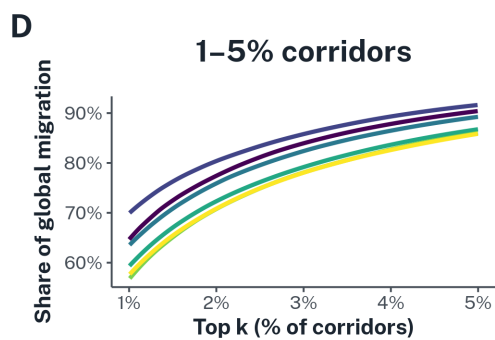
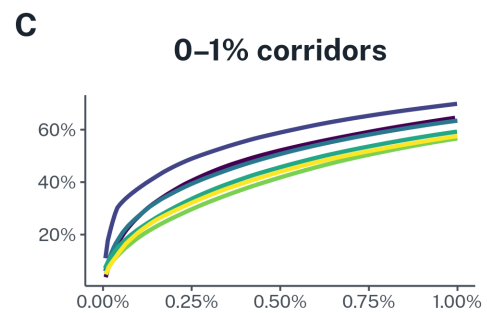
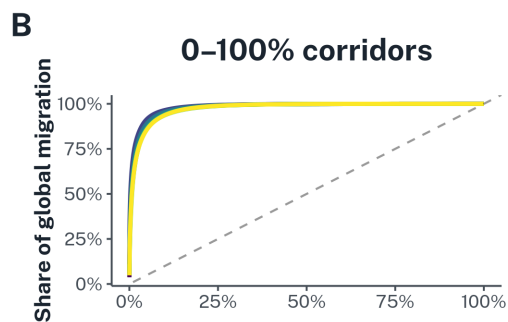
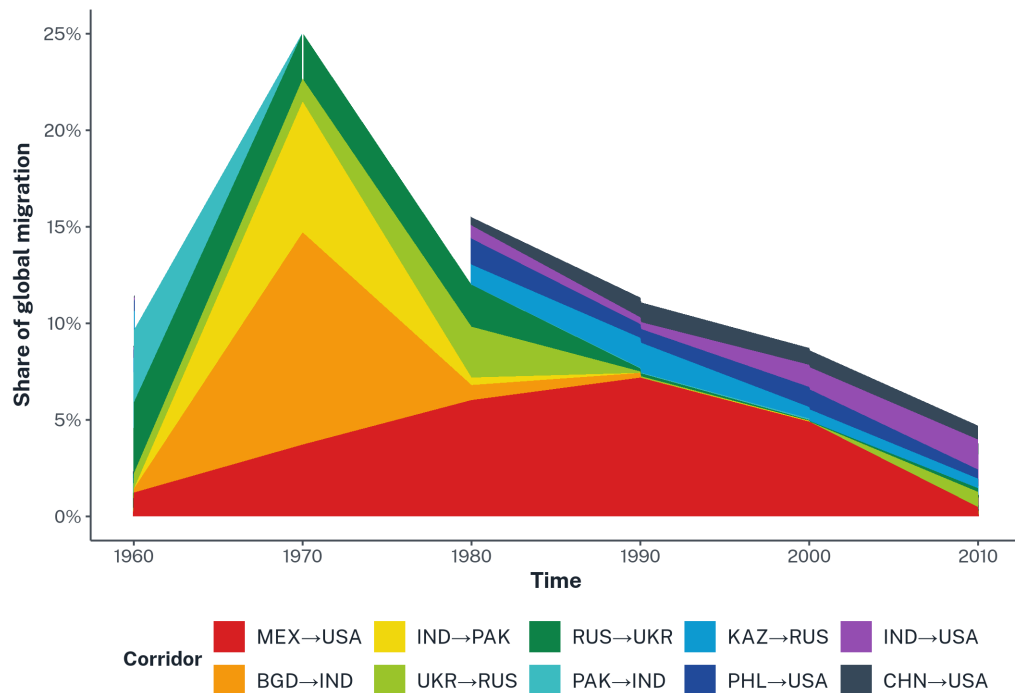


Figure A.33: Hill diversity of the global migration system, 1960–2020, using minimisation–open flow estimates. Panel A reports the Hill numbers for selected q values over time. Panel B reports the entire Hill profiles of each period, shown as differently-coloured lines.

A Share of global migration in the top 10 corridors (ranked by total flows)



Period

- 1960–1970
- 1970–1980
- 1980–1990
- 1990–2000
- 2000–2010
- 2010–2020

Figure A.34: Evolution of the dominance of the most important corridors in the global migration system, 1960–2020, using minimisation–open flow estimates.

Changes in global migration diversity under observed vs stable corridor set

Diversity change in observed system vs counterfactual with no corridor turnover.

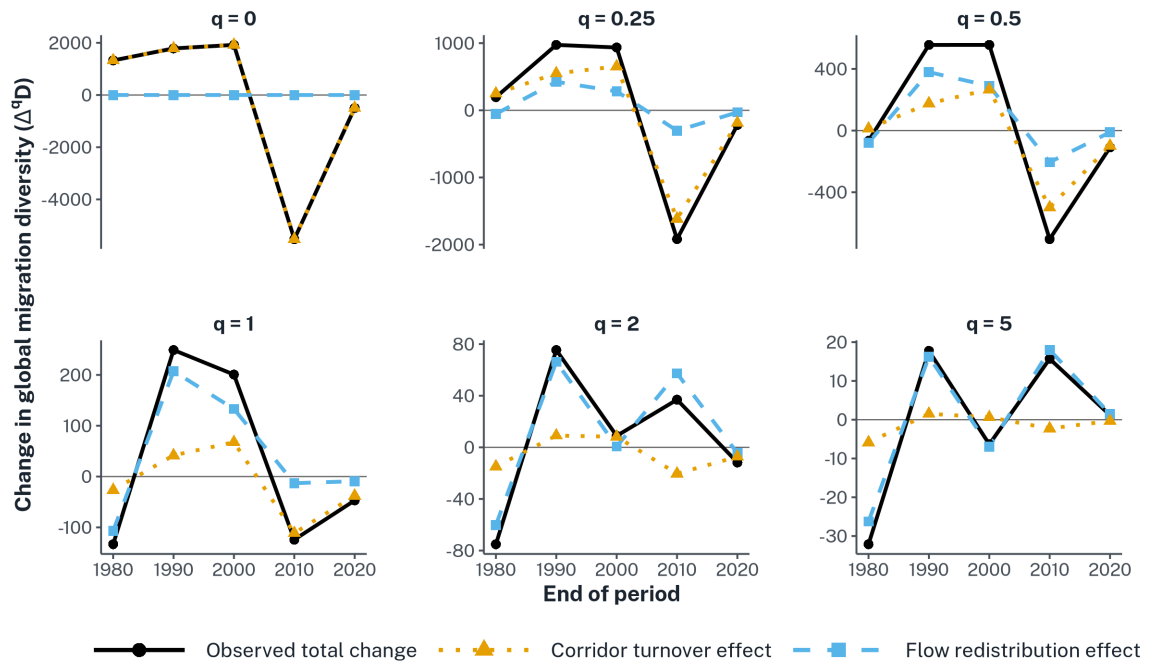


Figure A.35: Comparison of changes in global migration diversity through a stable-support analysis, using minimisation–open flow estimates.

Within- and between-country sources of migration diversity

Partitioning of system-level Hill diversity (qD) into alpha and beta components.

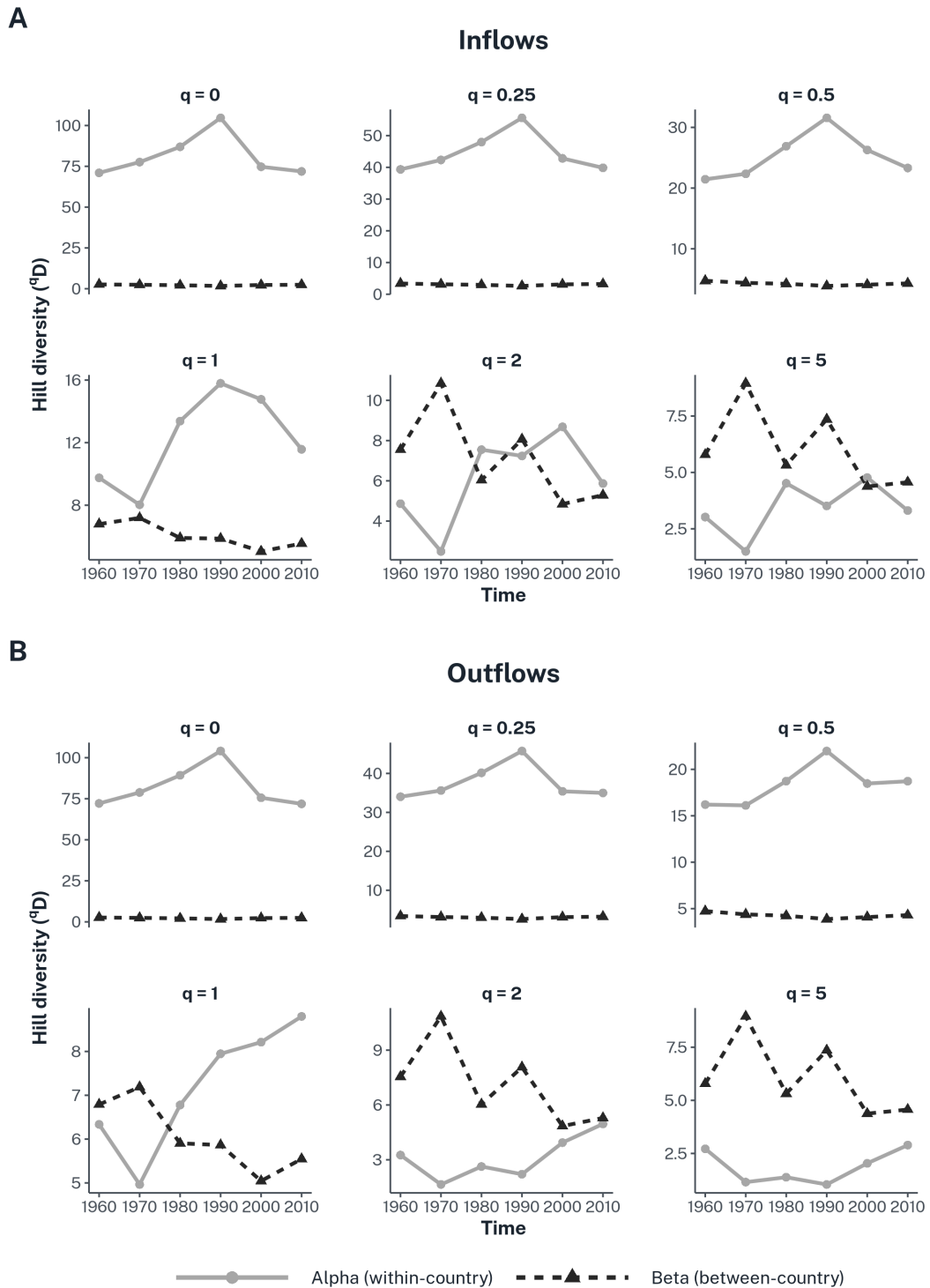
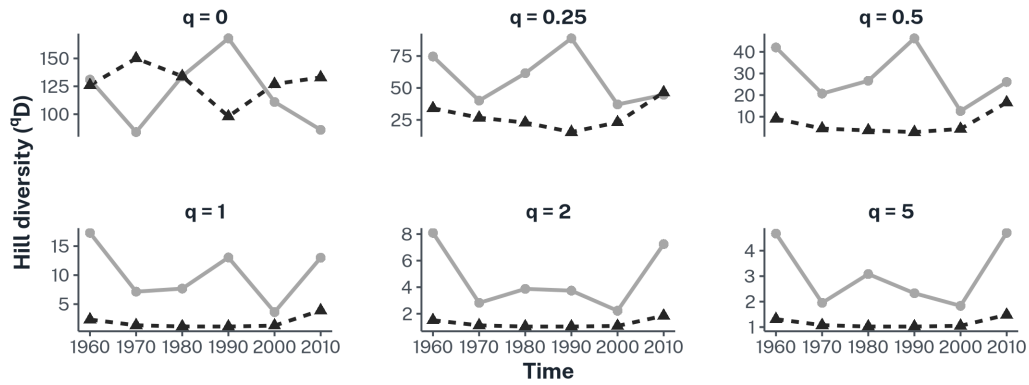


Figure A.36: Partitioning of global migration diversity into within- and between-country components, using minimisation–open flow estimates.

Diversity of migration inflows and outflows for selected countries

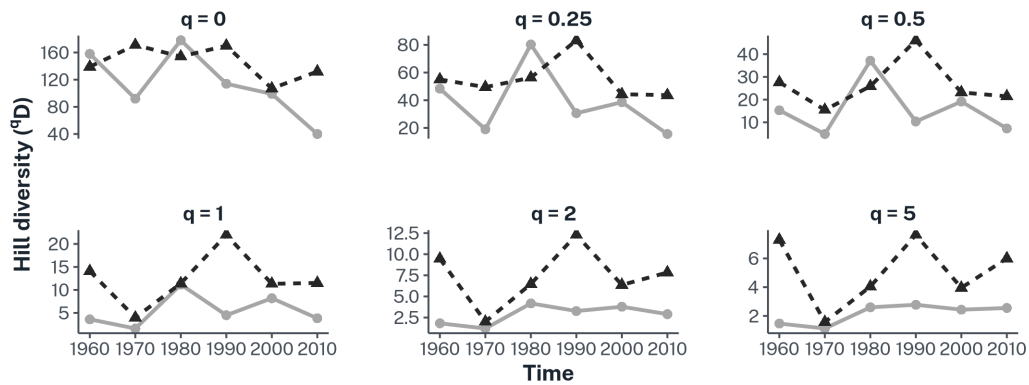
A

Mexico



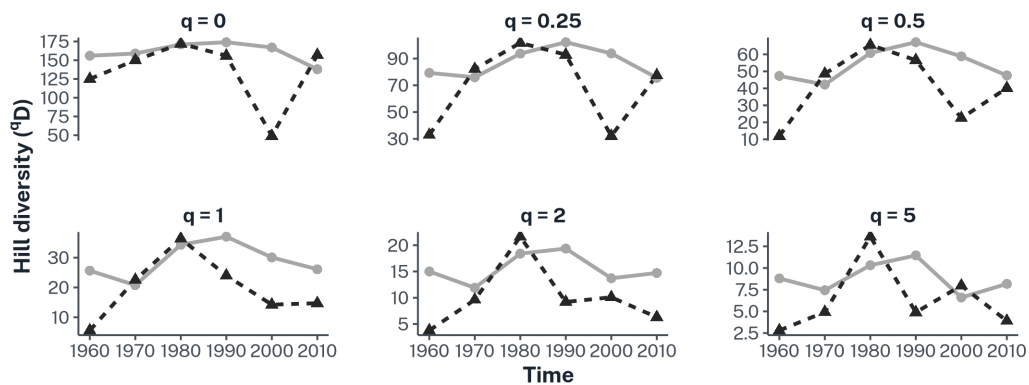
B

India



C

Italy



Flow type —●— Inflows - -▲- - Outflows

Figure A.37: Hill diversity of migration inflows and outflows for selected countries, 1960–2020, using minimisation–open flow estimates.

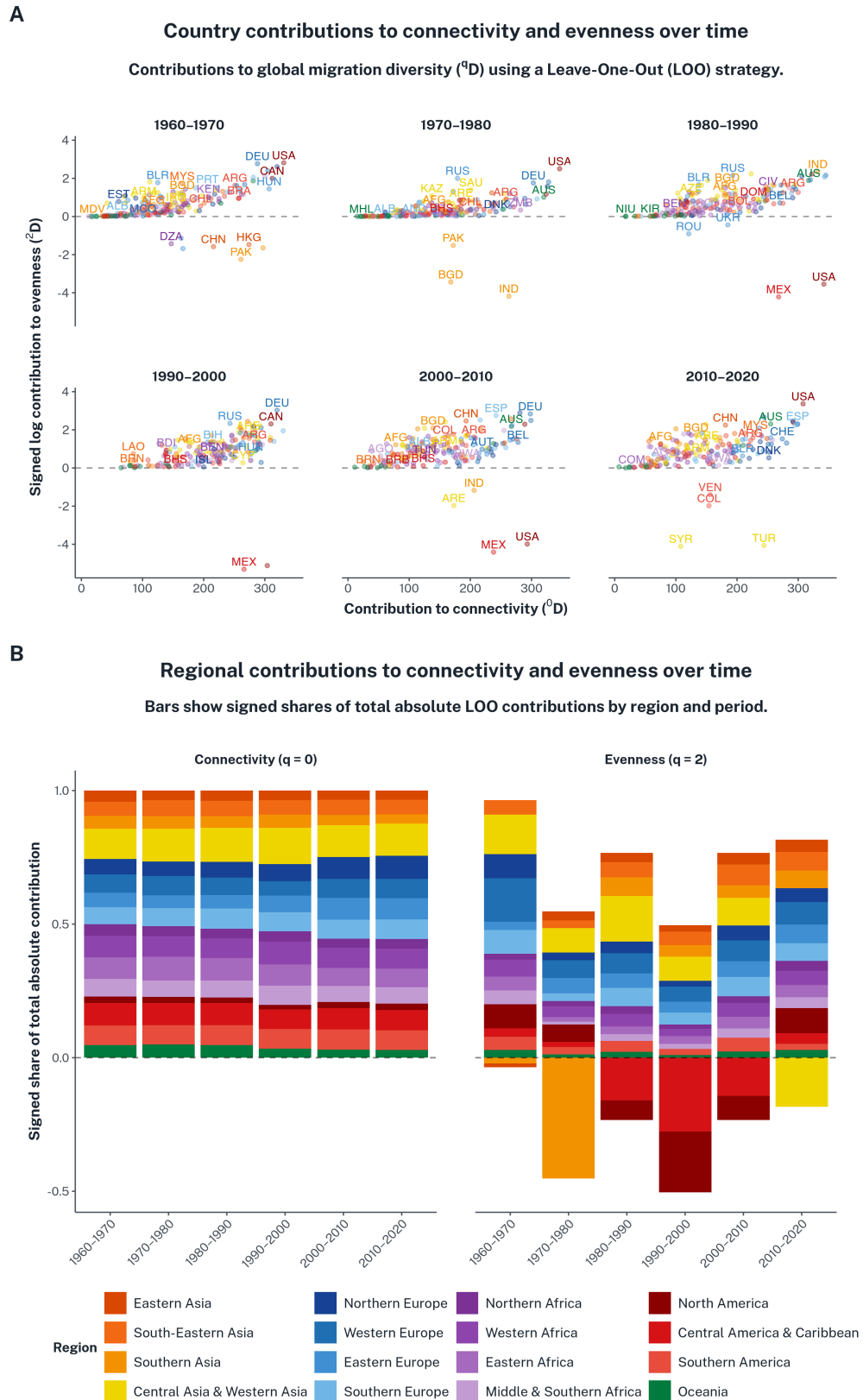


Figure A.38: Country- and region-level contributions to connectivity and evenness over time, using minimisation–open flow estimates.

B | Appendix to Chapter 4

B.1 Data Availability Statement

The migration flow data used in this chapter come from the Human Migration Database (HMigD), compiled and released by Dańko (2023), as detailed in Dańko et al. (2024). These data are available for academic use and can be accessed through the HMigD online platform: https://maciej-jan-danko.shinyapps.io/HMigD_Shiny_App_I. Additional variables used in the explanatory models are drawn from publicly available sources:

- (i) From the CEPII's Gravity database (Conte et al., 2021): geographical distance, measured as distance between capitals (`distcap`); contiguity (`contig`); shared official language (`comlang_off`); shared geopolitical history (`sibling_ever`); shared EU membership (`eu_both`).
- (ii) From the World Bank (World Bank, 2025a): GDP per capita, Purchasing Power Parity (PPP) in current international US dollars (\$; `NY.GDP.PCAP.PP.CD`).
- (iii) From the United Nations Department of Economic and Social Affairs (UNDESA; 2019, 2015): bilateral migrant stock in the year 2000.

As these data are produced and maintained by third parties, they cannot be uploaded directly to a public repository. However, the GitHub repository that accompanies the chapter, available at <https://github.com/mmorellini/spatial-mig-tests>, documents how each dataset can be accessed, along with all relevant links. Further, the GitHub repository includes the full set of processing steps, analysis scripts, and simulation code, enabling full replication of the regression models, simulations, and spatial evaluation metrics used in the chapter.

B.2 List of Countries in the Sample and Corresponding UN Macro-Regions

Table B.1: List of countries in the sample and corresponding ISO 3166-1 codes (alpha-2 and alpha-3) and UN world region labels (United Nations Statistics Division, 2024).

Country	ISO alpha-2	ISO alpha-3	UN Region
Austria	AT	AUT	Western Europe
Belgium	BE	BEL	Western Europe
Bulgaria	BG	BGR	Eastern Europe
Switzerland	CH	CHE	Western Europe
Cyprus	CY	CYP	Southern Europe
Czechia	CZ	CZE	Eastern Europe
Germany	DE	DEU	Western Europe
Denmark	DK	DNK	Northern Europe
Estonia	EE	EST	Northern Europe
Spain	ES	ESP	Southern Europe
Finland	FI	FIN	Northern Europe
France	FR	FRA	Western Europe
Greece	GR	GRC	Southern Europe
Croatia	HR	HRV	Southern Europe
Hungary	HU	HUN	Eastern Europe
Ireland	IE	IRL	Northern Europe
Iceland	IS	ISL	Northern Europe
Italy	IT	ITA	Southern Europe
Lithuania	LT	LTU	Northern Europe
Luxembourg	LU	LUX	Western Europe
Latvia	LV	LVA	Northern Europe
Malta	MT	MLT	Southern Europe
Netherlands	NL	NLD	Western Europe
Norway	NO	NOR	Northern Europe
Poland	PL	POL	Eastern Europe
Portugal	PT	PRT	Southern Europe
Romania	RO	ROU	Eastern Europe
Sweden	SE	SWE	Northern Europe
Slovenia	SI	SVN	Southern Europe
Slovakia	SK	SVK	Eastern Europe
United Kingdom	UK	GBR	Northern Europe

B.3 PPML Regressions with Varying Dependent Variables

Table B.2: PPML regressions of the baseline model with alternative dependent variables (migration flow quantiles). Alternative dependent variables were drawn from the HMigD posterior distribution (Daňko, 2023; Daňko et al., 2024) using different quantiles (q.).

Estimated migration flow:	5 th q. (0.05)	25 th q. (0.25)	50 th q. (0.5)	75 th q. (0.75)	95 th q. (0.95)
<i>Predictor variables:</i>					
– None (baseline specification)					
<i>Fixed effects:</i>					
Origin country	✓	✓	✓	✓	✓
Destination country	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓
<i>Fit statistics:</i>					
Observations	18,600	18,600	18,600	18,600	18,600
Squared correlation	0.55201	0.53452	0.52252	0.51068	0.49428
Adjusted pseudo-R ²	0.73536	0.72880	0.72398	0.71882	0.71117
Notes: Baseline specification includes only origin, destination, and year fixed effects. Adjusted pseudo-R ² is McFadden's R ² .					

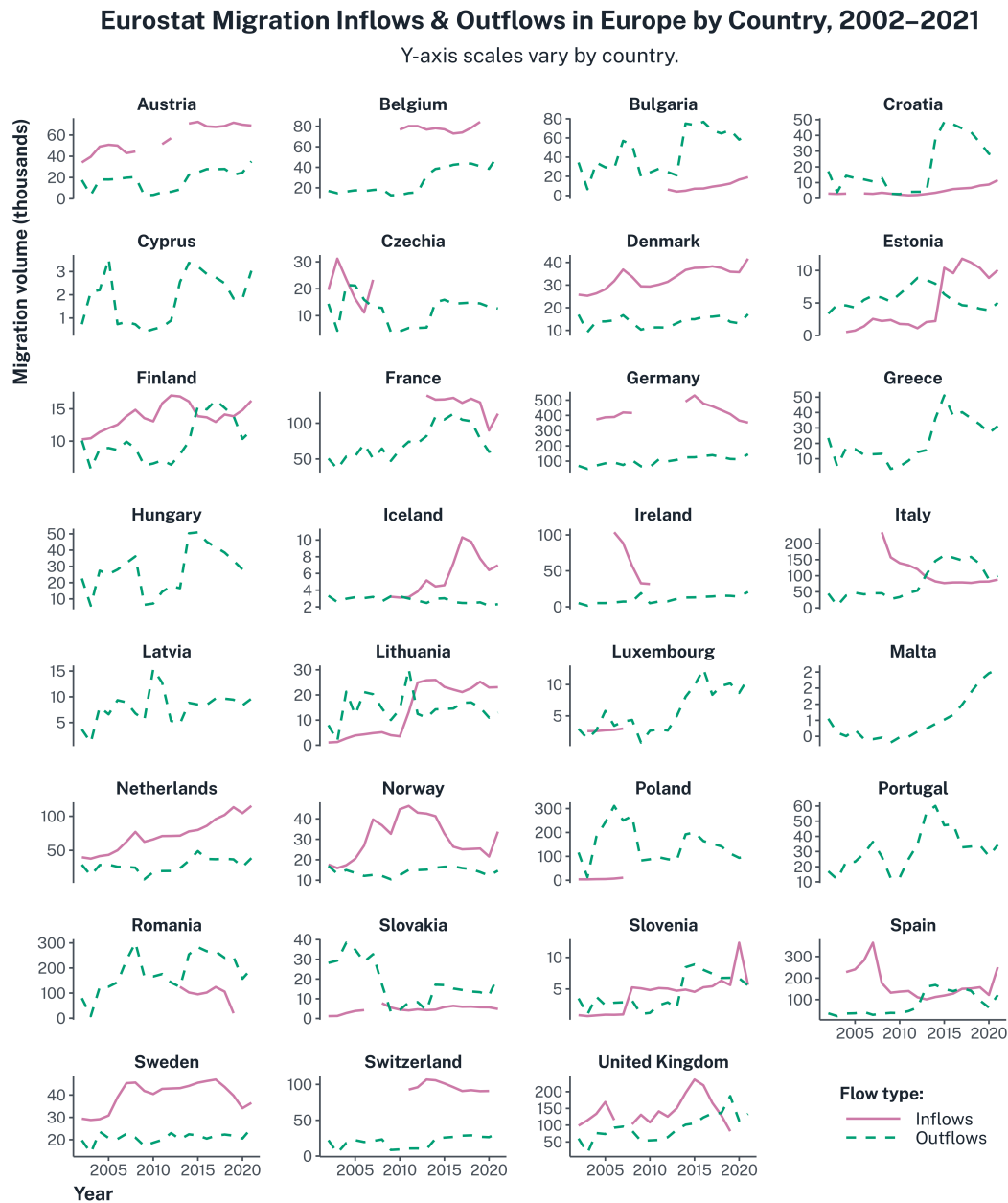
Table B.3: PPML regressions of the gravity model with alternative dependent variables (migration flows). Alternative dependent variables were drawn from the HMigD posterior distribution (Daňko, 2023; Daňko et al., 2024) using different quantiles (q.).

Estimated migration flow:	5 th q. (0.05)	25 th q. (0.25)	50 th q. (0.5)	75 th q. (0.75)	95 th q. (0.95)
<i>Predictor variables:</i>					
Distance (log)	−0.8286*** (0.1235)	−0.8095*** (0.1180)	−0.7955*** (0.1145)	−0.7812*** (0.1114)	−0.7608*** (0.1071)
Population ratio (lag, log)	−0.0629 (0.2920)	−0.0503 (0.2831)	−0.0454 (0.2756)	−0.0430 (0.2671)	−0.0431 (0.2536)
<i>Fixed effects:</i>					
Origin country	✓	✓	✓	✓	✓
Destination country	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓
<i>Fit statistics:</i>					
Observations	18,600	18,600	18,600	18,600	18,600
Squared correlation	0.61652	0.59943	0.58673	0.57345	0.55421
Adjusted pseudo-R ²	0.77594	0.77003	0.76556	0.76063	0.75315
Within Adjusted pseudo-R ²	0.15335	0.15204	0.15064	0.14871	0.14537
<i>Notes: Standard errors (in parentheses) clustered by origin, destination, and year.</i> <i>Adjusted pseudo-R² is McFadden's R². Significance codes: †: 0.1, *: 0.05, **: 0.01, ***: 0.001.</i>					

Table B.4: PPML regressions of the MST model with alternative dependent variables (migration flow quantiles). Alternative dependent variables were drawn from the HMigD posterior distribution (Daňko, 2023; Daňko et al., 2024) using different quantiles (q.).

Estimated migration flow:	5 th q. (0.05)	25 th q. (0.25)	50 th q. (0.5)	75 th q. (0.75)	95 th q. (0.95)
<i>Predictor variables:</i>					
Distance (log)	−0.2141 [†] (0.1203)	−0.2035 (0.1240)	−0.1959 (0.1266)	−0.1883 (0.1292)	−0.1772 (0.1320)
Population ratio (lag, log)	0.0891 (0.2897)	0.0944 (0.2881)	0.0930 (0.2845)	0.0881 (0.2786)	0.0723 (0.2655)
Contiguity	−0.1674 (0.1527)	−0.1630 (0.1527)	−0.1630 (0.1538)	−0.1661 (0.1558)	−0.1724 (0.1599)
Common official language	0.4802** (0.1843)	0.4967* (0.1937)	0.5117* (0.1996)	0.5301** (0.2052)	0.5606** (0.2109)
GDP ratio (lag, log)	0.5703 (0.6727)	0.4926 (0.6666)	0.4369 (0.6576)	0.3800 (0.6467)	0.3025 (0.6220)
Migrant stock (log)	0.4054*** (0.0604)	0.3959*** (0.0589)	0.3893*** (0.0578)	0.3828*** (0.0569)	0.3734*** (0.0558)
Shared geopolitical history	0.8658* (0.4072)	0.9419* (0.3993)	0.9911* (0.3918)	1.037** (0.3830)	1.086** (0.3660)
Shared EU membership (lag)	0.0208 (0.1294)	0.0137 (0.1257)	0.0034 (0.1251)	−0.0115 (0.1280)	−0.0421 (0.1416)
<i>Fixed effects:</i>					
Origin country	✓	✓	✓	✓	✓
Destination country	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓
<i>Fit statistics:</i>					
Observations	18,600	18,600	18,600	18,600	18,600
Squared correlation	0.67214	0.65837	0.64782	0.63658	0.61990
Adjusted pseudo-R ²	0.82812	0.82508	0.82276	0.82010	0.81574
Within adjusted pseudo-R ²	0.35052	0.35503	0.35789	0.36019	0.36208
<i>Notes: Standard errors (in parentheses) clustered by origin, destination, and year. Adjusted pseudo-R² is McFadden's R². Significance codes: †: 0.1, *: 0.05, **: 0.01, ***: 0.001.</i>					

B.4 Replication with Eurostat Data



Source: Eurostat (2023), Immigration by Age Group, Sex and Country of Previous Residence (MIGR_IMM5PRV).

Figure B.1: Eurostat estimates of aggregate migration inflows (in pink) and outflows (in green) in 31 European countries and over time.

Table B.5: Results of PPML regressions using Eurostat data.

Model:	Baseline	Gravity	MST
<i>Predictor variables:</i>			
Distance (log)		−0.9431*** (0.1468)	−0.2878* (0.1394)
Population ratio (lag, log)		−0.2212 (0.2451)	−0.0938 (0.1831)
Contiguity			−0.0065 (0.1305)
Common official language			0.7156*** (0.1762)
GDP ratio (lag, log)			−1.829** (0.6548)
Migrant stock (log)			0.3375*** (0.0616)
Shared geopolitical history			0.8608 (0.5542)
Shared EU membership (lag)			0.0660 (0.1973)
<i>Fixed effects:</i>			
Origin country	✓	✓	✓
Destination country	✓	✓	✓
Year	✓	✓	✓
<i>Fit statistics:</i>			
Observations	10,364	10,364	10,364
Squared correlation	0.55603	0.63634	0.65660
Adjusted pseudo-R ²	0.74695	0.79818	0.83895
Within adjusted pseudo-R ²		0.20265	0.36375
Ramsey RESET	0.0046	0.2047	0.1910
<i>Notes: Standard errors (in parentheses) clustered by origin, destination, and year.</i> <i>Adjusted pseudo-R² is McFadden's R². RESET p values computed following Santos Silva and Tenreyro (2006). Signif. codes: †: 0.1, *: 0.05, **: 0.01, ***: 0.001.</i>			

B.5 Replication with HMigD (Excluding 2019–2021)

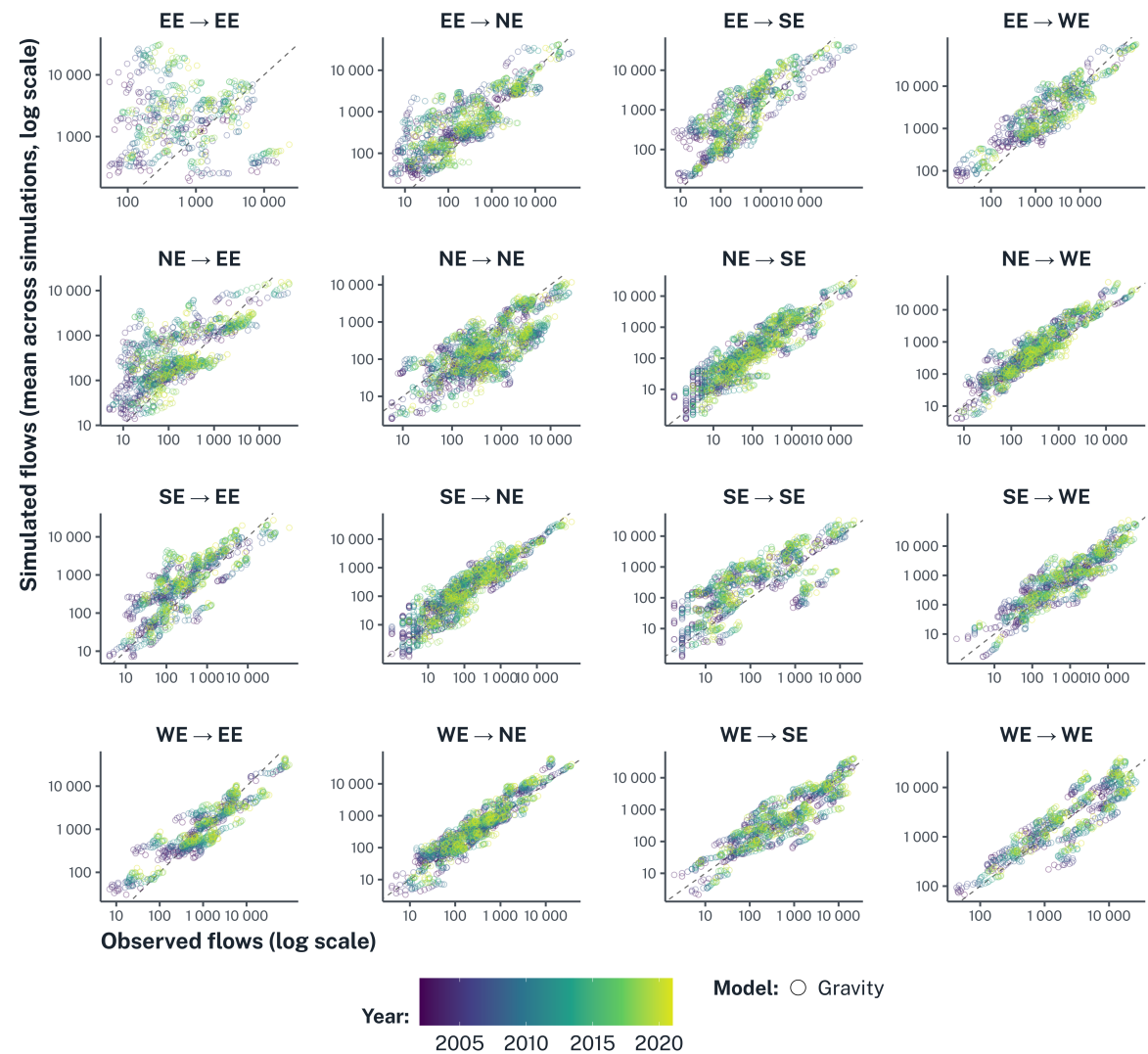
Table B.6: Results of PPML regressions with HMigD data, excluding years affected by the COVID-19 pandemic (2019–2021).

Model:	Baseline	Gravity	MST
<i>Predictor variables:</i>			
Distance (log)		−0.8418*** (0.1125)	−0.2452* (0.1168)
Population ratio (lag, log)		−0.0685 (0.2733)	0.0434 (0.2665)
Contiguity			−0.1572 (0.1567)
Common official language			0.5277** (0.1917)
GDP ratio (lag, log)			1.166 [†] (0.6272)
Migrant stock (log)			0.3980*** (0.0591)
Shared geopolitical history			0.8904* (0.3650)
Shared EU membership (lag)			0.1260 (0.1751)
<i>Fixed effects:</i>			
Origin country	✓	✓	✓
Destination country	✓	✓	✓
Year	✓	✓	✓
<i>Fit statistics:</i>			
Observations	15,810	15,810	15,810
Squared correlation	0.52137	0.59950	0.66666
Adjusted pseudo-R ²	0.72198	0.76804	0.82780
Within adjusted pseudo-R ²		0.16568	0.38062
Ramsey RESET	0.0287	0.1584	0.1674
<i>Notes: Standard errors (in parentheses) clustered by origin, destination, and year.</i> <i>Adjusted pseudo-R² is McFadden's R². RESET p values computed following Santos Silva and Tenreyro (2006). Signif. codes: †: 0.1, *: 0.05, **: 0.01, ***: 0.001.</i>			

B.6 Scatterplots with Observed *vs* Simulated Flows

Gravity: Simulated vs Observed Flows by Corridor-Year and Region Pair

Each point is a corridor-year observation; dashed lines indicate perfect simulation.

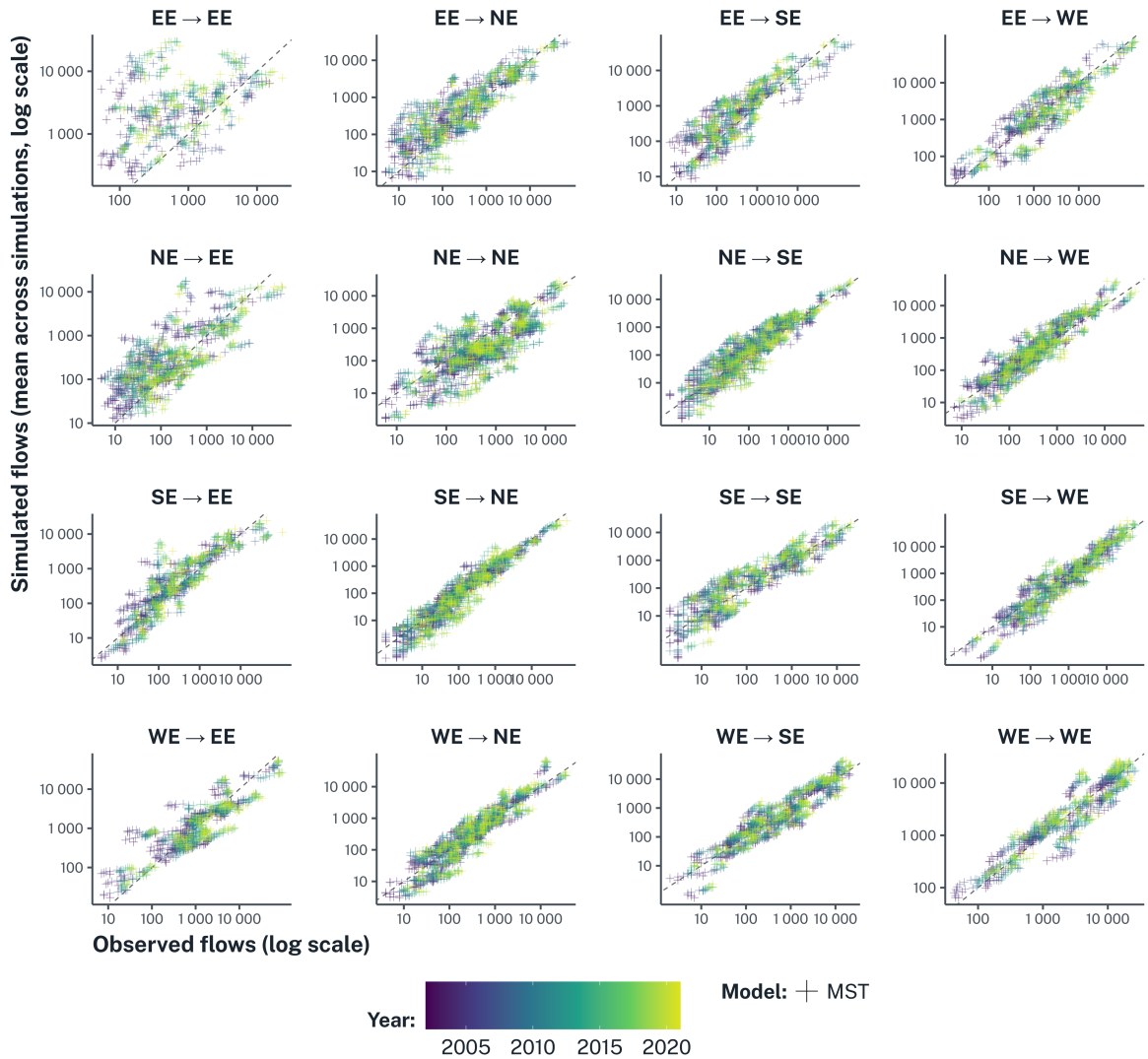


Region labels: EE = Eastern Europe, NE = Northern Europe, SE = Southern Europe, WE = Western Europe.

Figure B.2: Comparison of the gravity-simulated and observed migration flows by region-level corridor over time. For each corridor-year combination, the points compare the observed flows with the mean of 10,000 simulated flows under the gravity model (circles). The diagonal indicates perfect simulation: simulated flows are equal to the observed ones. The scatterplots are grouped by the regions to which the 31 countries belong, as origins or destinations: Eastern Europe (EE), Northern Europe (NE), Southern Europe (SE), and Western Europe (WE).

MST: Simulated vs Observed Flows by Corridor-Year and Region Pair

Each point is a corridor-year observation; dashed lines indicate perfect simulation.



Region labels: EE = Eastern Europe, NE = Northern Europe, SE = Southern Europe, WE = Western Europe.

Figure B.3: Comparison of the MST-simulated and observed migration flows by corridor, year, and time. For each corridor-year combination, the points compare the observed flows with the mean of 10,000 simulated flows under the MST model (crosses). The diagonal indicates perfect simulation: the simulated flows are equal to the observed ones. The scatterplots are grouped by the regions to which the 31 countries belong, as origins or destinations: Eastern Europe (EE), Northern Europe (NE), Southern Europe (SE), and Western Europe (WE).

Simulated vs Observed Flows by Corridor–Year and Model

Corridor–year observations for gravity (circles) and MST (triangles) in grey. Selected corridors highlighted and coloured in blue (gravity) and orange (MST).

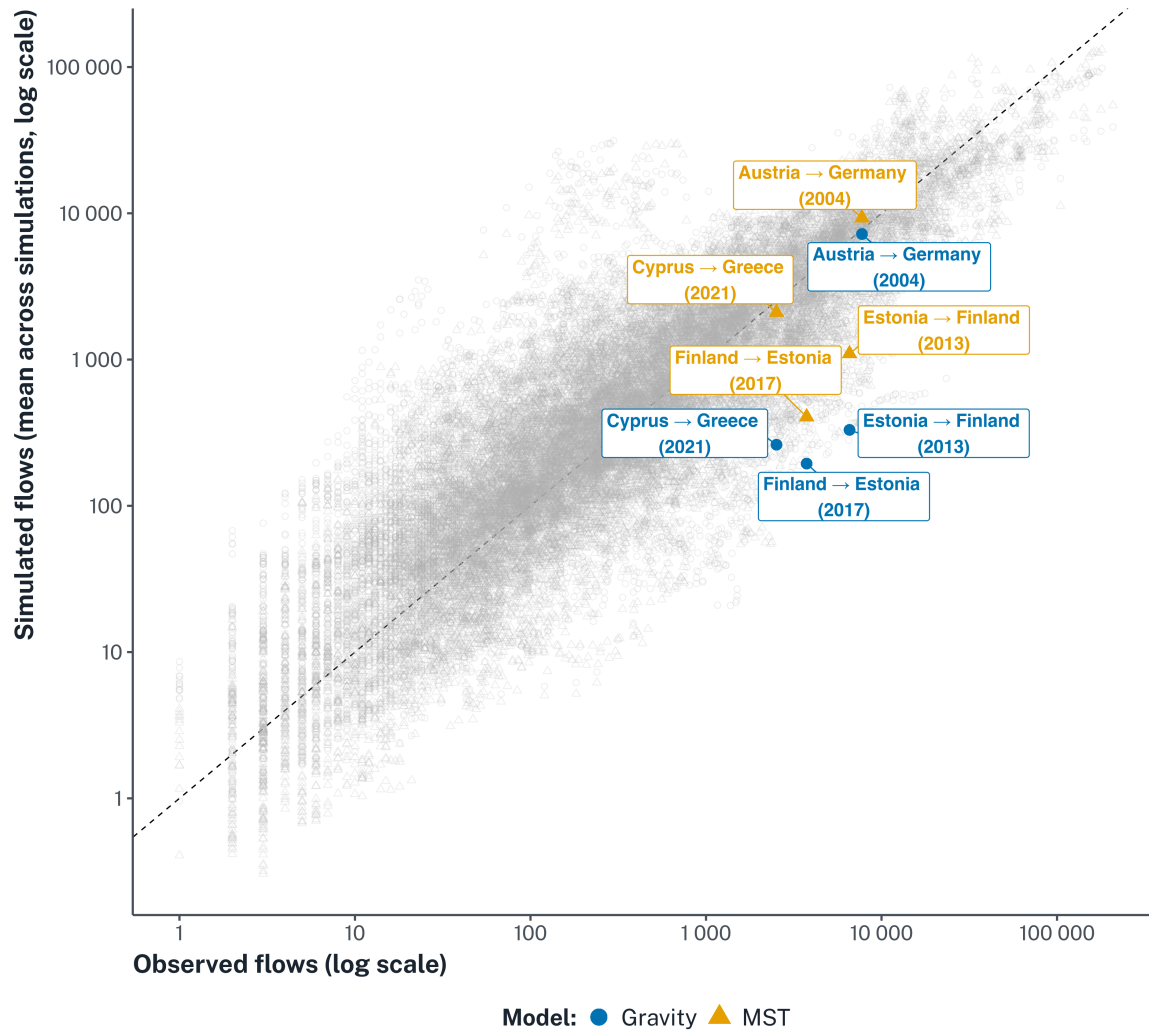


Figure B.4: Comparison of simulated and observed migration flows by corridor, year, and model. For each corridor-year combination, the points compare the observed flows with the mean of 10,000 simulated flows under two explanatory models, gravity (dots, highlighted in blue) and MST (triangles, highlighted in orange). The diagonal indicates perfect simulation: the simulated flows equal the observed ones. Four corridor-year combinations are highlighted: (1) flows from Austria to Germany in 2004; (2) flows from Cyprus to Greece in 2021; (3) flows from Estonia to Finland in 2013; and (4) flows from Finland to Estonia in 2017. These four cases were chosen to illustrate where and when simulated flows may deviate from the observed ones, illustrating model-specific strengths and weaknesses.

C | Appendix to Chapter 5

C.1 Data Availability Statement

The data used in this chapter come from multiple publicly available sources. Data on Erasmus mobility were obtained from the European Commission (2021), while data on degree mobility were compiled jointly by the UNESCO Institute for Statistics (UIS), the Organisation for Economic Cooperation and Development (OECD), and Eurostat (UIS, 2023). Together, these cover all international student moves for academic purposes within 29 European countries in the calendar year 2019. Additional variables used in the explanatory models are drawn from open-access sources:

- (i) From the CEPII's Gravity database (Conte et al., 2021): geographical distance, measured as distance between capitals (`distcap`); contiguity (`contig`); shared official language (`comlang_off`).
- (ii) From the World Bank (2025a, 2025b): GDP per capita, Purchasing Power Parity (PPP) in current international US dollars (`$; NY.GDP.PCAP.PP.CD`) and population size (`SP.POP.TOTL`) in 2018.
- (iii) From the Human Migration Database (HMigD; Daňko, 2023; Daňko et al., 2024): country-to-country migration flows in 2018.

As these data are produced and maintained by third parties, they cannot be uploaded directly to a public repository. However, the full set of data processing scripts, regression models, simulation code, and corridor-level evaluation metrics will be made available via a dedicated online repository upon acceptance of the article version of the chapter, enabling full replication of the study.

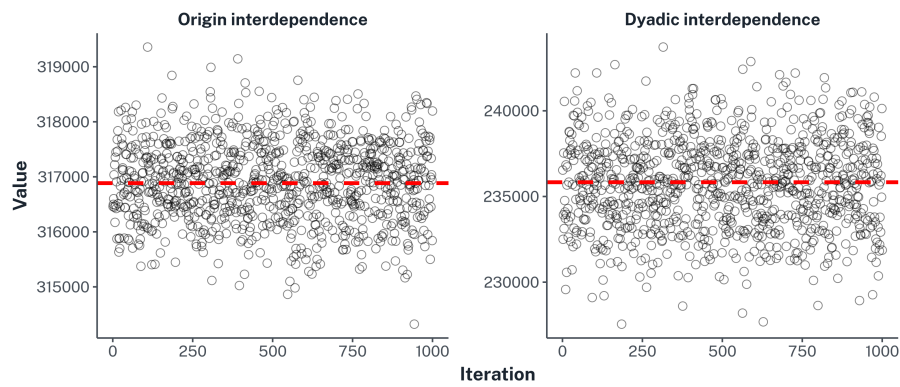
C.2 MCMC Convergence Diagnostics of ELMos with Interdependence Parameters (Models 3–4)

Trace plots of MCMC estimation: Degree sample, Model 3

Observed values in red dashed lines.

A

Trace plots by iteration



B

Density plots of simulated values

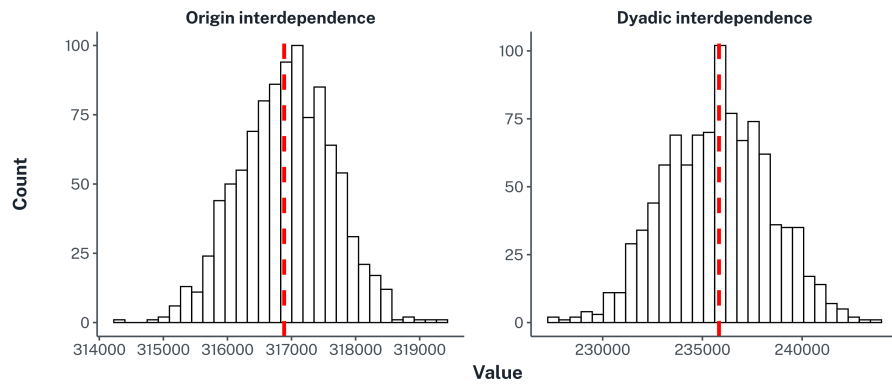


Figure C.1: Diagnostic plots of Model 3 for degree mobility. Panel A shows trace plots of simulated statistics across MCMC iterations, while Panel B displays the corresponding distributions of simulated values. Red dashed lines denote observed network statistics. The absence of systematic trends in the trace plots and the close alignment between simulated and observed values indicate good mixing and convergence of the Markov chains, suggesting that the estimation has reached stationarity and that the model provides an adequate fit to the observed network.

Diagnostics of MCMC estimation: Degree sample, Model 4

Observed values in red dashed lines.

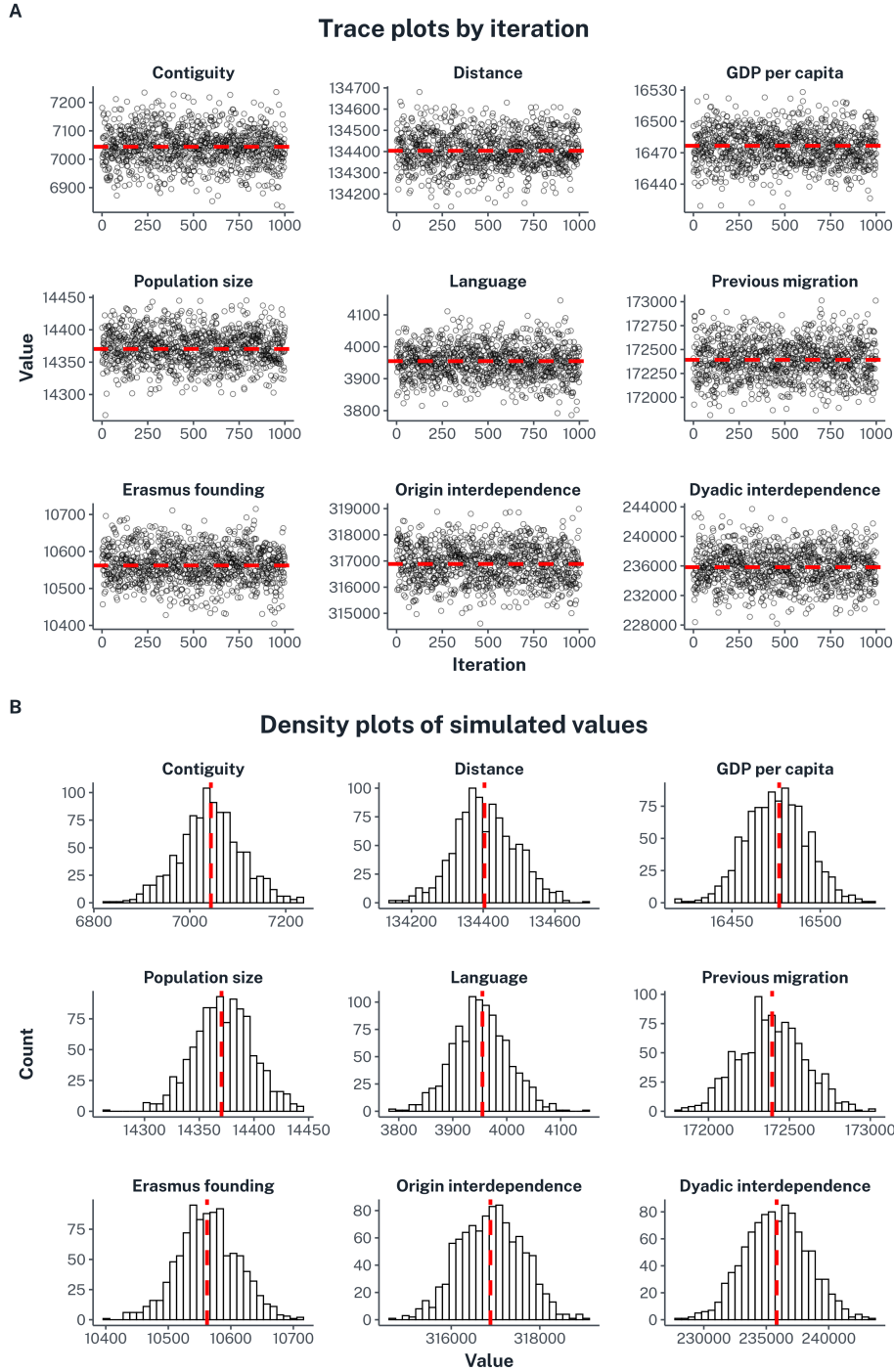


Figure C.2: Diagnostic plots of Model 4 for degree mobility. Panel A shows trace plots of simulated statistics across MCMC iterations, while Panel B displays the corresponding distributions of simulated values. Red dashed lines denote observed network statistics. The absence of systematic trends in the trace plots and the close alignment between simulated and observed values indicate good mixing and convergence of the Markov chains, suggesting that the estimation has reached stationarity and that the model provides an adequate fit to the observed network.

Diagnostics of MCMC estimation: Erasmus sample, Model 3

Observed values in red dashed lines.

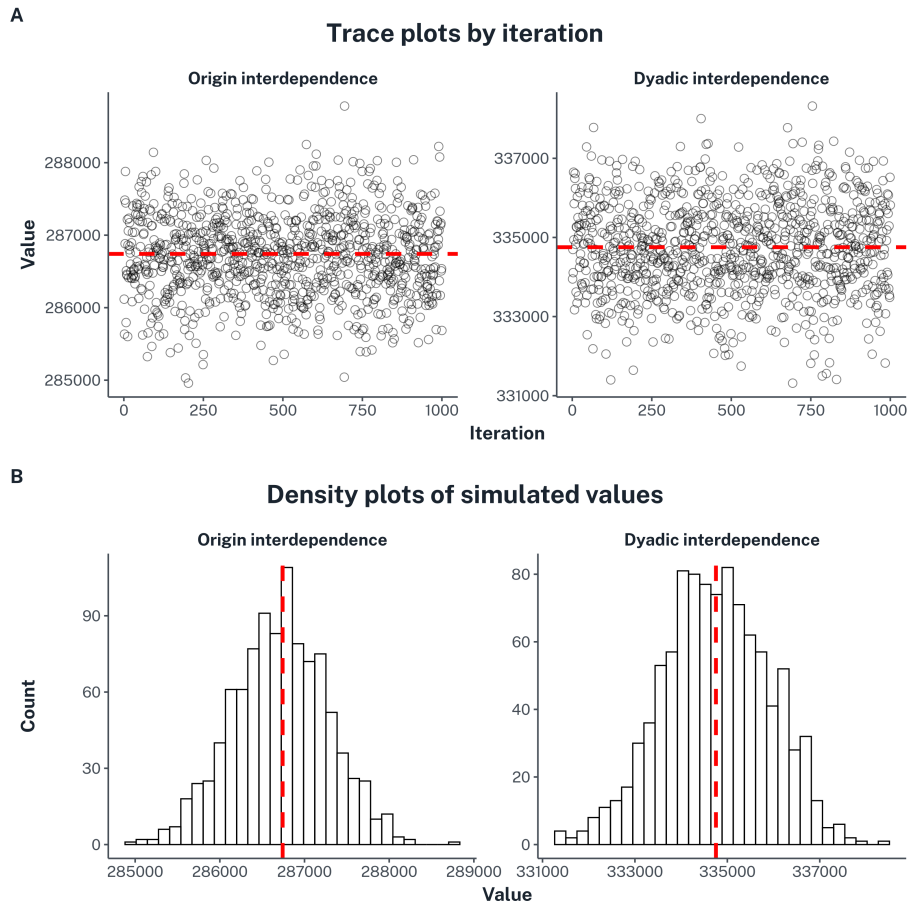


Figure C.3: Diagnostic plots of Model 3 for Erasmus mobility. Panel A shows trace plots of simulated statistics across MCMC iterations, while Panel B displays the corresponding distributions of simulated values. Red dashed lines denote observed network statistics. The absence of systematic trends in the trace plots and the close alignment between simulated and observed values indicate good mixing and convergence of the Markov chains, suggesting that the estimation has reached stationarity and that the model provides an adequate fit to the observed network.

Diagnostics of MCMC estimation: Erasmus sample, Model 4

Observed values in red dashed lines.

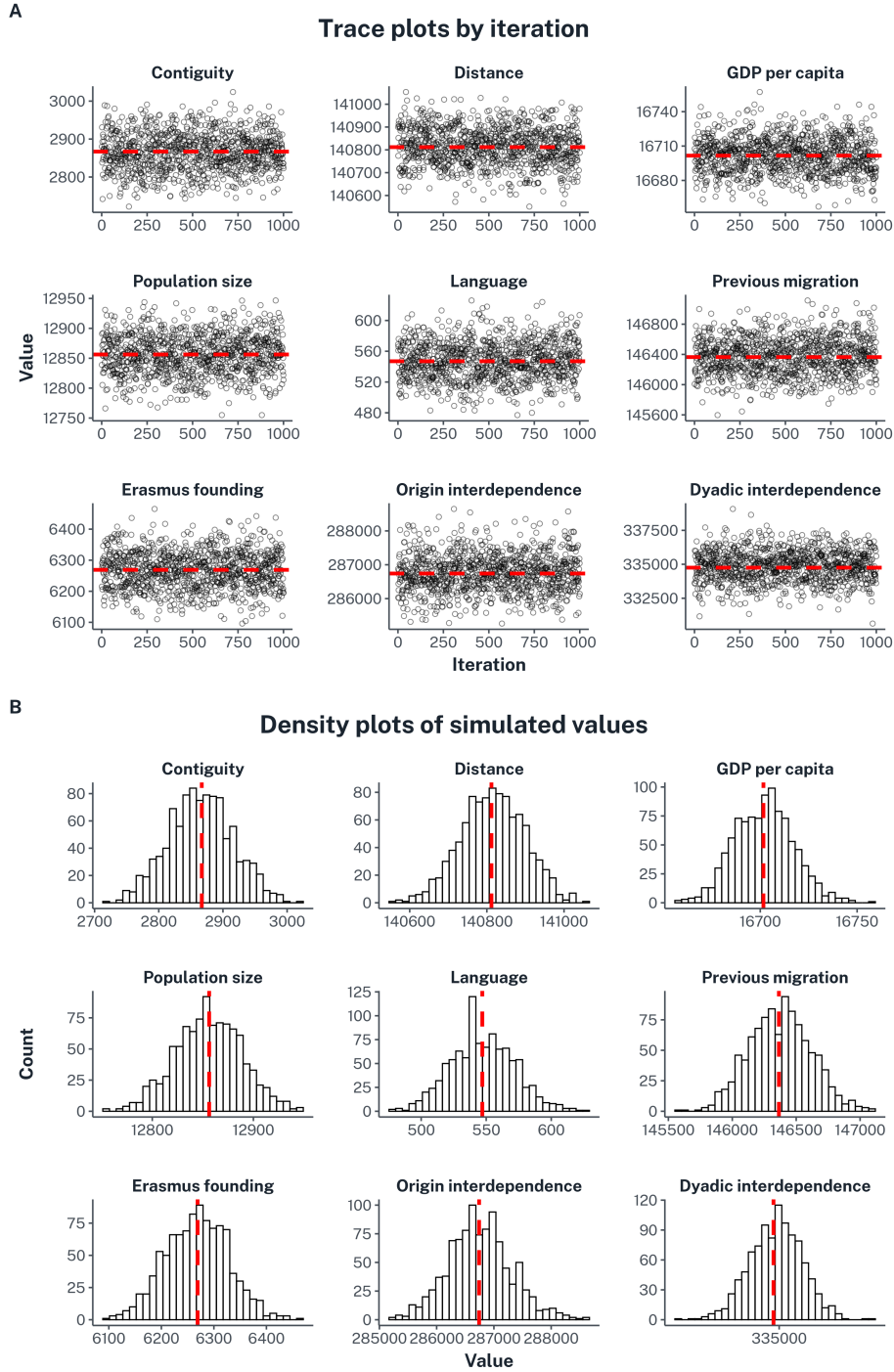


Figure C.4: Diagnostic plots of Model 4 for Erasmus mobility. Panel A shows trace plots of simulated statistics across MCMC iterations, while Panel B displays the corresponding distributions of simulated values. Red dashed lines denote observed network statistics. The absence of systematic trends in the trace plots and the close alignment between simulated and observed values indicate good mixing and convergence of the Markov chains, suggesting that the estimation has reached stationarity and that the model provides an adequate fit to the observed network.

C.3 Prediction Bias by Corridor for All Models

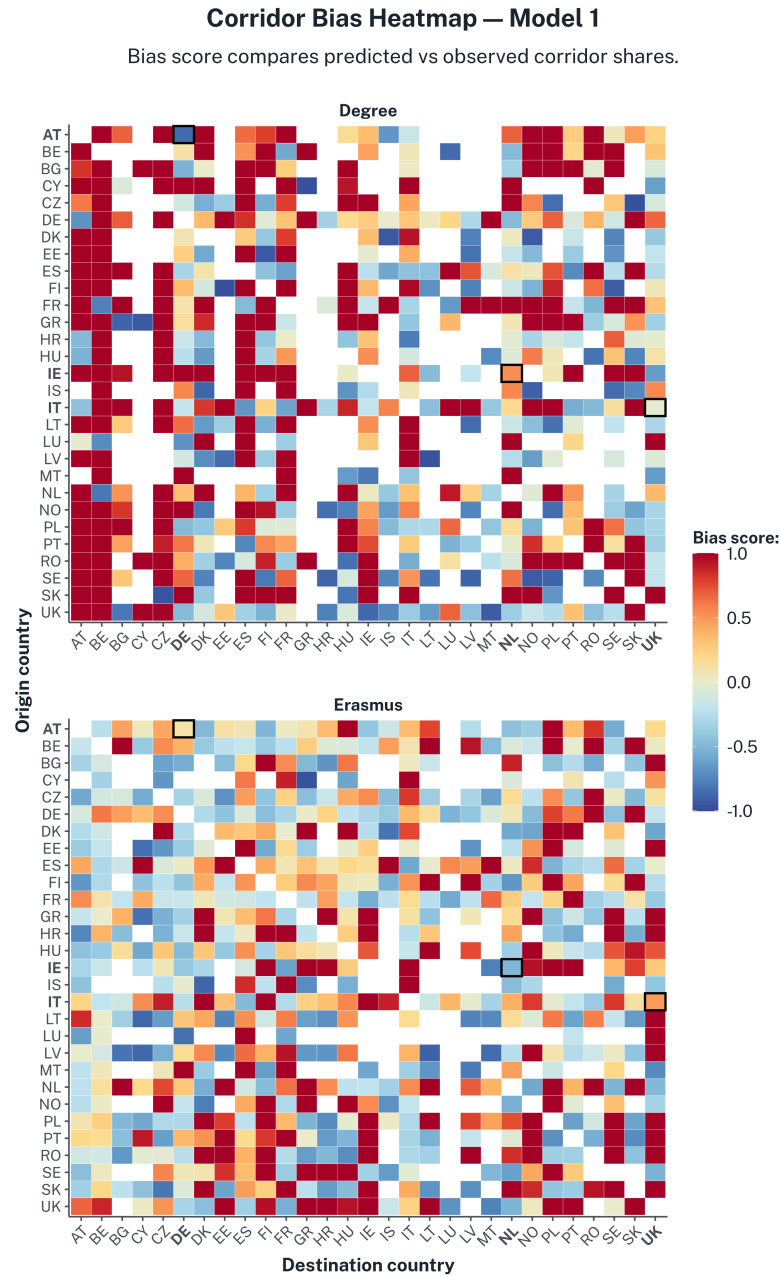


Figure C.5: Prediction bias of Model 1 for all migration corridors in the sample. Each cell displays the relative deviation between predicted and observed migration flows for a given origin–destination corridor. Blue indicates under-prediction while red indicates over-prediction; the shading reflects the magnitude of the bias metric, with less intense hues indicating better prediction and more intense hues worse prediction. Three exemplary corridors (Austria to Germany, Ireland to the Netherlands, and Italy to the UK) are highlighted through bold lettering and borders.

Corridor Bias Heatmap — Model 2

Bias score compares predicted vs observed corridor shares.

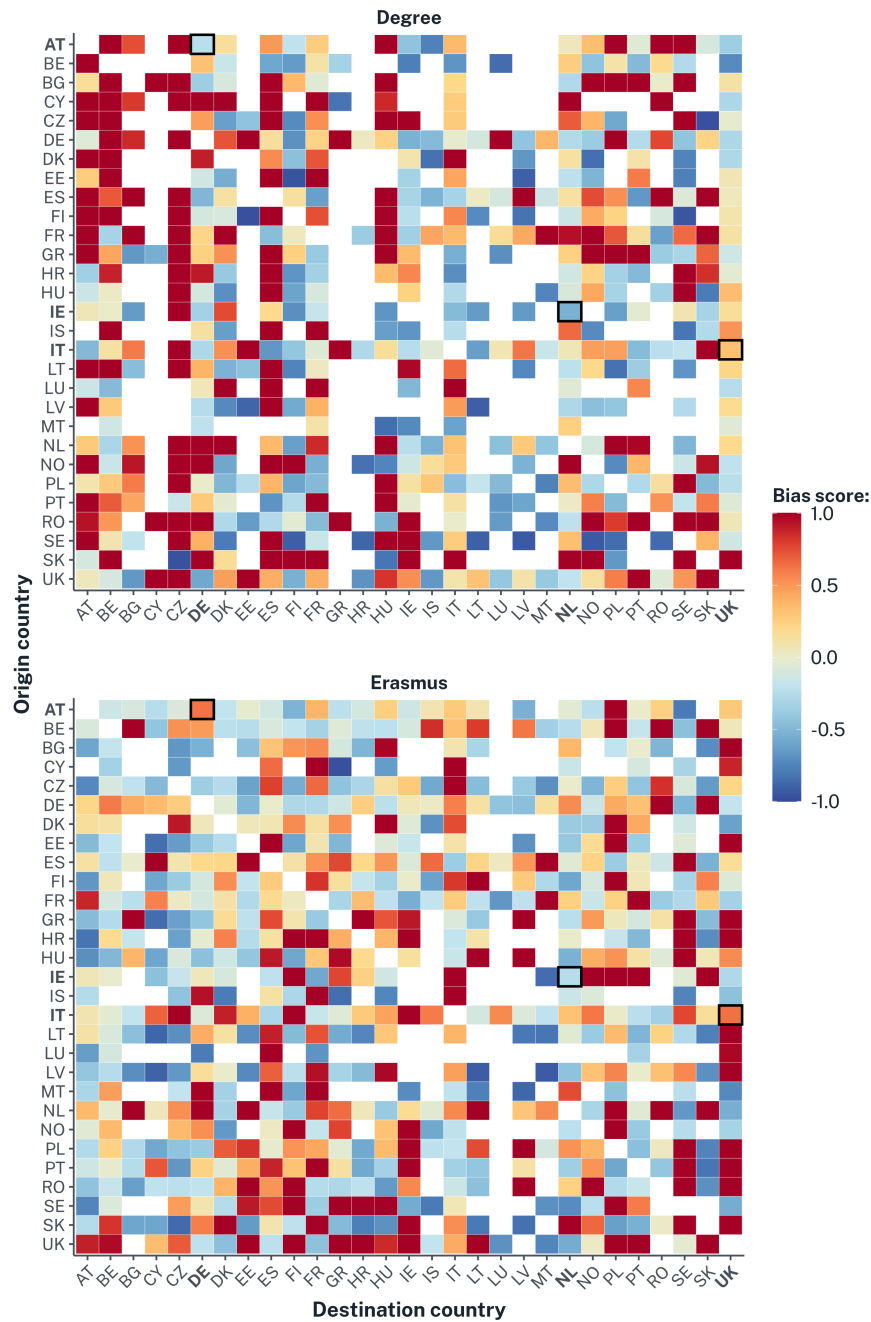


Figure C.6: Prediction bias of Model 2 for all migration corridors in the sample. Each cell displays the relative deviation between predicted and observed migration flows for a given origin–destination corridor. Blue indicates under-prediction while red indicates over-prediction; the shading reflects the magnitude of the bias metric, with less intense hues indicating better prediction and more intense hues worse prediction. Three exemplary corridors (Austria to Germany, Ireland to the Netherlands, and Italy to the UK) are highlighted through bold lettering and borders.

Corridor Bias Heatmap — Model 3

Bias score compares predicted vs observed corridor shares.

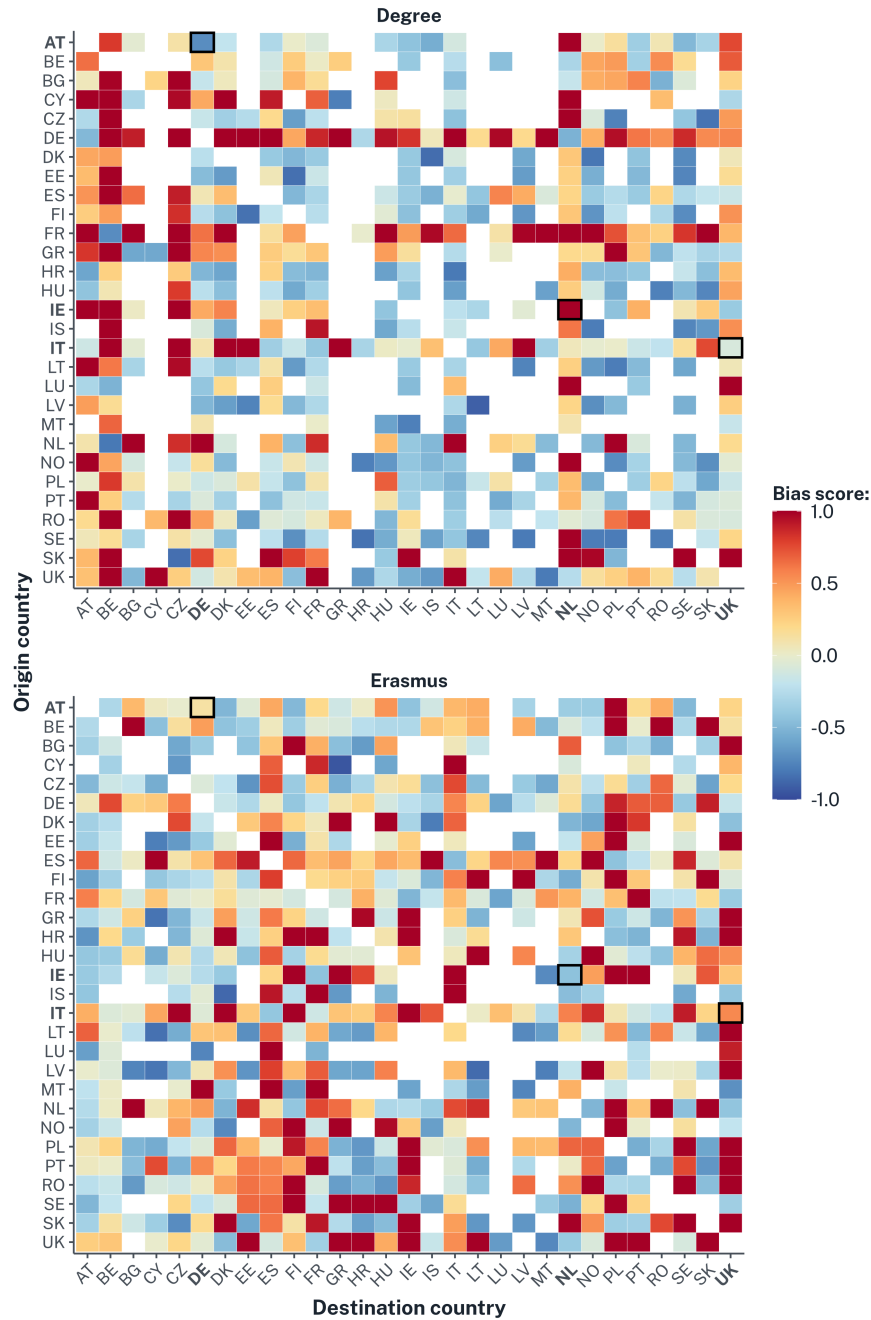


Figure C.7: Prediction bias of Model 3 for all migration corridors in the sample. Each cell displays the relative deviation between predicted and observed migration flows for a given origin–destination corridor. Blue indicates under-prediction while red indicates over-prediction; the shading reflects the magnitude of the bias metric, with less intense hues indicating better prediction and more intense hues worse prediction. Three exemplary corridors (Austria to Germany, Ireland to the Netherlands, and Italy to the UK) are highlighted through bold lettering and borders.

C.4 Corridor-Level Simulation Accuracy for All Models

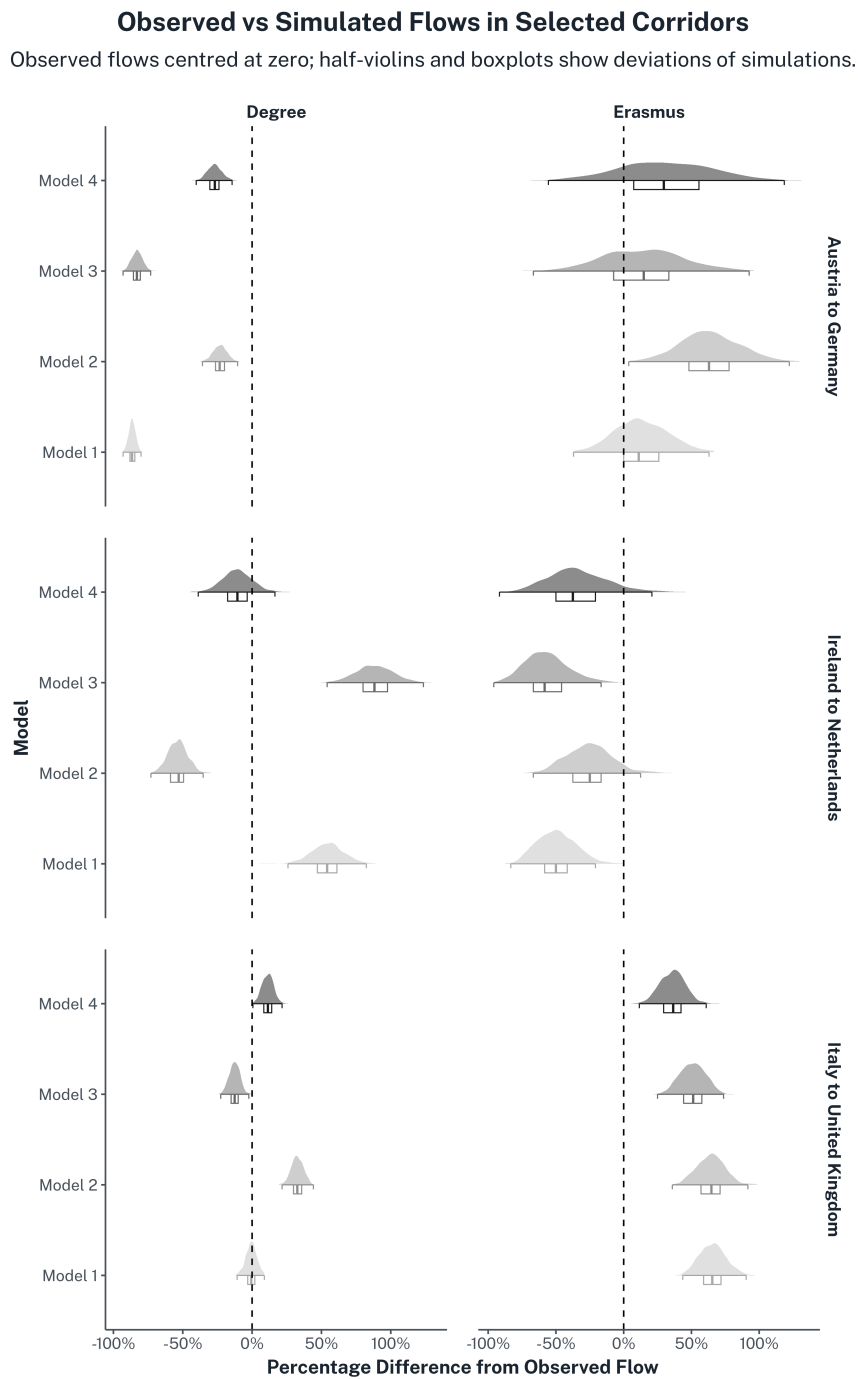


Figure C.8: Performance comparison across all models for selected migration corridors. For three exemplary corridors (Austria to Germany, Ireland to the Netherlands, Italy to the UK) and by ISM type, 1,000 counterfactual networks were simulated per model; simulated corridor flows are shown relative to observed flows.