SciencesPo
SCHOOL OF PUBLIC AFFAIRS

PUBLIC POLICY MASTER THESIS

May 2021

Benefits on-site or to take away:

A panel data analysis on how infrastructures shape emigration

Tobias Hillenbrand

Master's Thesis supervised by Golvine de Rochambeau

Second member of the Jury: Pierre-Philippe Combes

Master in Public Policy
Economics and Public Policy

Abstract

In light of upward-pointing migration trends, policymakers continue to suggest socioeconomic development as a means to reduce the number of international immigrants despite plenty of contradicting evidence. However, uncertainty remains about what exactly drives the emigration rates of lower-income countries on their path to economic prosperity. As the role of infrastructures is underexplored, I contribute to the investigation of the infrastructure-emigration relationship by proposing a novel classification that distinguishes between an infrastructure type with benefits tied to locations (place-based) and another type with benefits tied to persons (people-based). Due to differences in the "portability" of benefits, I hypothesize that the place-based type affects emigration negatively, whereas people-based infrastructures should have a positive impact. I compile a panel dataset with data from UN DESA, the Global Competitiveness Index, and other sources, that includes 120 countries and covers a period from 2005 to 2020. My hypotheses are tested through the application of different regression analysis techniques, among them are fixed effects and instrumental variable regressions. The findings suggest that place-based infrastructures are likely to affect emigration negatively, while the picture for education – representing the people-based type – is more mixed. The results are of high political relevance and encourage further research (e. g., using individual-level data) to create a deeper understanding of the patterns of cross-border movements.

Key words

migration, emigration, infrastructures, place-based, people-based, development

Table of content

List of abbreviations

BLUE	Best Linear Unbiased Estimator
CUR	Country of Usual Residence
e. g.	exempli gratia
et al.	et alii
etc.	et cetera
fe	fixed effects
GCI	Global Competitiveness Index
GDP	Gross Domestic Product
ICT	Information and Communication Technology
i. e.	id est
IOM	International Organization for Migration
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
p.	page
PEI	People-based Infrastructures
PLI	Place-based Infrastructures
PPP	Purchasing Power Parity
re	random effects
TFP	Total Factor Productivity
TSLS	Two-Stage-Least-Squares
UN	United Nations
UN DESA	United Nations Department of Economic and Social Affairs
WEF	World Economic Forum
vs.	versus

List of tables

List of figures

Why should I read this research?

Migration may be as old as the history of mankind and yet it ranks among the top issues of political agendas around the globe. Even though the share of international migrants in the overall world population remains on a modest level, trends are pointing upward, and regional differences are large. While cross-border mobility brings about many benefits, the past years have also uncovered the powerful potential of the issue to create deep divisions both within and between societies. As a result, policymakers are increasingly seeking ways to better steer migration flows.

Steering, however, requires a sound understanding of the matter, and many statements around the narrative of "fighting the root causes of migration", which frequently implies accelerating the socioeconomic development of origin countries, seem to be guided by an important misconception. Overall, the alleviation of poverty and the creation of better-paid jobs, as desirable as that might be, cannot be expected to drive down the number of emigrants. In fact, the mobility-transition curve that represents the empirical relationship between emigration and economic development, follows the shape of an inverted U. Hence, economic advancement of low-income countries is likely to be associated with *higher* emigration rates.

This empirical observation departs from the standard neoclassical doctrine that the relationship between income differences across countries and migration costs should determine the international mobility of people. While different explanations for this puzzle have been put forward, much uncertainty remains with respect to the role of specific development components. Therefore, it remains unclear if all kinds of development investments necessarily increase emigration rates in low-income societies.

In this context, the relationship between infrastructures and emigration – whose analysis has so far been mostly limited to cases of within-country mobility – is investigated. Infrastructures create benefits for people through the channels of incomes and amenities, but – as I argue – the benefits differ in their degree of "portability". Based on this difference, I develop a novel classification of infrastructures into a place-based (e. g., transport and electricity infrastructure) and a people-based (e. g., education infrastructure) type, create a theoretical framework to derive hypotheses regarding their effects on emigration and test these hypotheses with a variety of regression analyses using an extensive dataset.

Therefore, my work contributes to a better understanding of the relationship between development and migration in general as well as infrastructures and emigration in particular. I present results that encourage complementary research and identify promising future research approaches. Besides, this paper provides insights for policymakers with an awareness of the complexity of the migration issue who are seeking for differentiated responses to today's challenges.

1 Introduction

In his famous book "Exodus – How migration is changing our world", the Oxford Professor Paul Collier (2013, p. 12) states that the issue of migration "has been politicized before it has been analyzed". Even though parts of Collier's analysis have been challenged by other leading scholars[1] of the field, there is an overwhelming consensus among migration researchers and international organizations that the public debate on this hot political topic is frequently detached from scientific evidence and tends to be dominated by alarmism, emotions and frequently biased perceptions (Biffl, 2019; de Haas et al., 2020; International Organization of Migration (IOM), 2019; Koser, 2016; Organisation for Economic Co-operation and Development (OECD), 2017). In the aftermath of the 2015 migration crisis, the potential of the migration issue to create deep divisions within and between societies was revealed and even jeopardized the future of the European project (Mechitishvili, 2021).

Even though the numbers of international migrants are on the rise, foreign-born people remain a small minority of the world population. As I will argue, there is good reason to assume that migration flows will continue to increase, which does not necessarily represent a burden on destination or sending countries but will certainly keep the issue very high on political agendas. Regardless of the normative assessment of migration, there is evidently a widely spread misconception with respect to its drivers, particularly in the context of migration and development. The "addressing the root causes of migration"-narrative (Clemens and Postel, 2018) does not only portray migration primarily as a threat for rich destination countries, but also uncovers a poor understanding of the role of socioeconomic development. The rationale to use economic advancement as a containment strategy for migration is rooted in the "standard" neoclassical approach that seeks to explain cross-border movements with a narrow focus on income disparities across countries. However, plenty of evidence on the so-called "mobility-transition curve" shows that, by and large, increasing income levels in lower-income countries are associated with *higher* rather than lower rates of emigration. This deviation of the empirics from the standard theoretical predictions constitutes a puzzle that has been addressed by multiple researchers but still entails a good amount of open questions (Clemens, 2020, 2014; Dao et al., 2018; Ortega and Peri, 2013).

For example, as many changes occur alongside economic development, it is difficult disentangle the effects of individual factors (Clemens, 2014). Hence, various components of socioeconomic development or large-scale investments in origin countries may impact migration patterns in very different ways, but be dominated by other factors (Clist and Restelli, 2020). Against this background, I focus on the role of infrastructures as a key element of development that received surprisingly little attention in international migration research so far. Moreover, no distinction has yet been made about different types of infrastructures depending on whether their benefits are tied to places (e. g., transport or electricity lines), or people (e. g., education infrastructure). I refer to them as place-based infrastructures (PLI) and people-based infrastructures (PEI) and argue that, given the differences in the "portability" of the benefits, they should incentivize migration in different ways. In the light of these considerations, I intend to address the following research question:

[1] For example, a critical review by Clemens and Sandfur (2014) was printed in the Foreign Affairs magazine.

What are the effects of place-based and people-based infrastructures on emigration rates?

To tackle the question, I proceed as follows: The second chapter provides an overview of the international migration subject. To this end, I define the key concepts, and outline migration trends and patterns in the migrant population. Moreover, I discuss risks and opportunities of cross-border movements, summarize the relationship between development and emigration as well as existing findings on the effects of infrastructures on migration.

In the third chapter I present a novel classification that could serve to better understand the infrastructure-emigration relationship. I also introduce a theoretical model that highlights the channels through which infrastructure types (dis)incentivize emigration. Based on these insights, I derive my hypotheses, which can then be tested.

Chapter 4 starts with developing an empirical framework that builds on key insights from the literature and my theoretical model and identifies the necessary data that is presented in the second part of this chapter. Migration flow data is obtained from the United Nations Department of Economic and Social Affairs' (UN DESA) database on migrant stocks, whereas Global Competitiveness Index (GCI) data from the World Economic Forum (WEF) is used for my infrastructure variables. The following statistical examinations in chapter 5 stretch from the investigation of bivariate associations to the conduction of a set of multivariate regression computations. I present different techniques for the analysis of panel data, but also examine the more sustained flows by building averages across the entire observation period. An instrumental variable approach is also included.

Chapter 6 discusses the findings critically and highlights a number of relevant limitations, while the final chapter summarizes the key points of this work, gives policy recommendations and points to meaningful open questions that require further attention.

2 State of research on migration and development

In this chapter, I provide an interdisciplinary overview of the literature on migration and development. In a first step, I clarify relevant definitions, before speaking about current and expected migration trends. I also discuss opportunities and risks of migration and summarize the literature on the relationships between development as well as infrastructures and emigration.

2.1 Relevant definitions in the migration context

To ensure terminological clarity, I begin by providing definitions for the key terms, which is particularly important in view of the oftentimes confusing public discussions on migration. First, I want to stress that I focus on *international* migration. Several differentiations are essential to bear in mind, such as the difference between migration (flow variable) and migrants (stock variable) (Inglis et al., 2019). The former can be relatively easily defined as the movement of people that involves crossing national borders (IOM, 2019).[2] The definition of a migrant on the other hand can depend on different characteristics whose relevance varies across regions. In European countries, for example, the discrepancy of citizenship and the country of usual residence (CUR) is

[2] Note however, that not any kind of border movement is regarded as a form of migration (e. g., commuting to work does not qualify). Only cross-border mobility with the purpose of changing one's country of usual residence are considered migratory movements (Inglis et al., 2019).

often crucial. The drawback of this approach is that persons who have never crossed any border can be categorized as migrants, if their parents do not hold the citizenship of the CUR (UN DESA, 2019a). In countries with a longer immigration history, it is rather the deviation of a person's birth place from their CUR that classifies someone as a migrant (Inglis et al., 2019; IOM, 2019a).

The "foreign-born" approach seems more consistent in combining the act of migration with the classification of migrants and is prevalent among international organizations that also provide the data for the empirical investigation below (IOM, 2019b; OECD, 2021; UN, 2021). Therefore, I adopt this definition.

Defining a migrant as a person whose CUR deviates from their country of birth leads to the need to clarify the meaning of CUR in a next step. The very general answer by the IOM's (2019b) Glossary on Migration is: "[T]he country in which a person has his or her usual or habitual residence". This definition is somewhat spelled out by the UN DESA, defining CUR as "[t]he country in which a person lives, that is to say, the country in which he or she has a place to live, where he or she normally spends the daily period of rest" (UN DESA, 1998, p. 92). Oftentimes a differentiation is made between long-term (a length of stay of at least twelve months) and short-term migrants (between three and twelve months) (IOM, 2019a; OECD, 2021; UN, 2021).

While I do not distinguish between migrants based on their legal status (e. g., regular vs. irregular migrants), the difference between voluntary and forced migration matters. Labor migration is the most prominent form of voluntary migration and defined by the IOM (2019b) as "the movement of persons from one state to another or within their own country of residence for the purpose of employment". Other forms of voluntary migration can range from student migrants to movements for "social reasons", i. e., moving to a partner or family reunions (Koser, 2016). Forced migration, on the other hand, is described as "a migratory movement, which, although the drivers can be diverse, involves force, compulsion, or coercion" IOM (2019b). Examples for forced migration can be (extreme) forms of violence, political persecution, or environmental factors. The affected people are often labelled as refugees although this term can also be defined in a narrower sense. Therefore, I prefer the more general term "humanitarian migrants" (de Haas et al., 2020, p. 31). Due to my research idea, I am interested in voluntary and particularly in labor migration. The line between forced and voluntary migration is often blurred though. While employers may exert a lot of pressure on employees to move to a foreign branch, even most forcefully displaced people face a choice of moving within or across national borders (Koser, 2016).

To recap briefly, I define (international) migrants as foreign-born persons and focus especially on labor migration. If a migrant is regarded as an immigrant (arriving person) or an emigrant (departing person) depends on whether the perspective of the receiving country or of the sending country is adopted. Given my interest in understanding why people leave, rather than where they go to, I usually take the perspective of the country of origin and thus speak about emigration and emigrants. I use the terms "sending country", "origin country" and "home country" as well as "receiving country", "destination country" and "host country" interchangeably.[3]

[3] A discussion on the different connotations of these expressions is provided by (de Haas et al., 2020), but these differences are not regarded essential for my research interest.

Having established a common base regarding the terminology, I proceed by giving an overview of past and current migration trends.

2.2 Migration trends

In recent years, migration has been a fiercely debate topic, particularly in European countries and the United States. In the aftermath of the so-called "European migration crisis", anti-immigrant parties have been on the rise in Europe, while Donald Trump's promise to build a border wall to Mexico was one of his campaign hits that might have secured him the presidency (de Haas et al., 2020; McCaskill, 2016; Mechitishvili, 2021). A summary of recent trends of migration can help understand and assess the salience of the matter. More importantly though, this overview of the volumes and directions of migration flows, the underlying drivers as well as the composition of the migrant population serves as the foundation for the migration model that is subsequently developed. In line with my research interest, I do not cover specifically trends of forced migration.

First, it is important to acknowledge the estimation of the World Migration Report 2020 that approximately 3.5 percent of the world population were international migrants in 2019, corresponding to an absolute number of 272 million people. While the overall percentage has remained subject to only small and gradual increases, the rise of migration has been much more pronounced in absolute numbers (IOM, 2019a).[4] Table 1 highlights the evolution of the numbers of international migrants in relative and absolute terms.

Table 1: Evolution of migration numbers

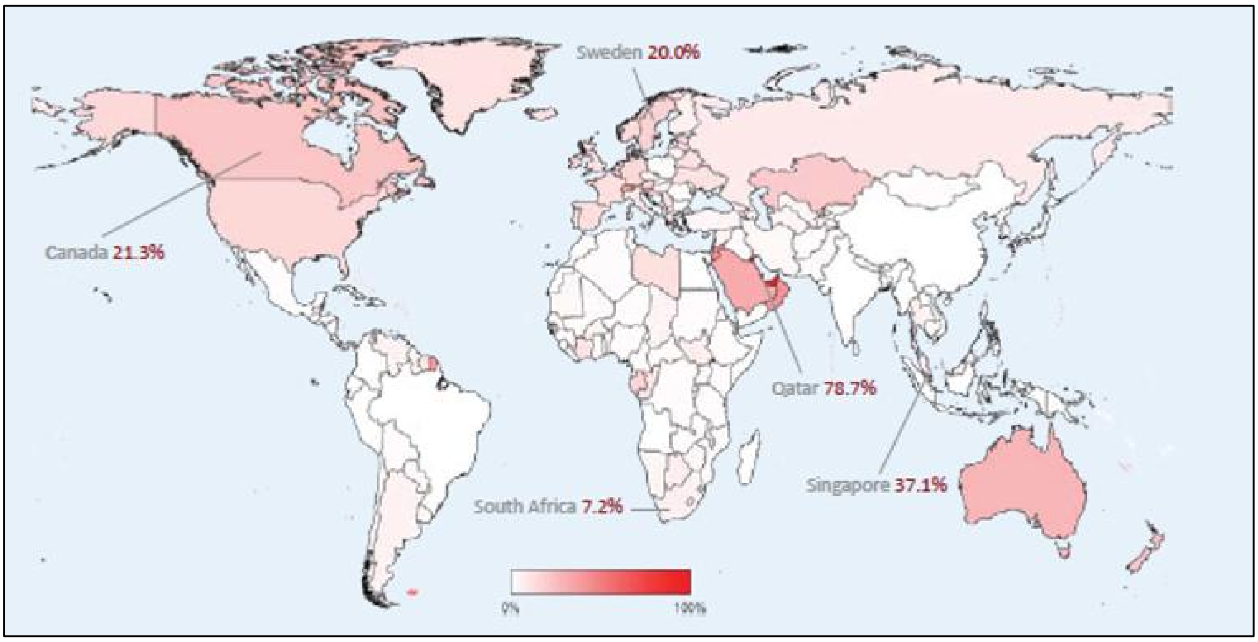
Year	Number of international migrants	International migrants as % of world population
2000	174 million	2.8
2010	220 million	3.2
2019	272 million	3.5

Source: IOM (2019a), based on data from UN DESA.

A more differentiated picture can be painted when looking beyond the global numbers and considering the large differences between regions and countries. Figures A-1 to A-6[5] highlight the magnitudes of these interregional differences by uncovering migration flows to, within and from the various global regions. Despite the extremely diverse picture, it is remarkable though that in all regions and all categories migration trends are at least slightly positive with the only exception being migration from Europe. The link between income levels and migration trends that the regional graphs already imply is confirmed in figure 1. The darker colors of most industrialized countries suggest a positive association between migration stocks and income levels (IOM, 2019a).

[4] De Haas (2007a) notes that today's levels are not necessarily unprecedented in relative terms. Before the First World War, the share of international migrants in the world population was at a comparable level.
[5] Note that all figures and tables that contain an "A" are printed in the appendix.

Figure 1: Share of migrants in the population by country



Source: IOM (2019a).

Overall, roughly two thirds of migrants live in high-income and almost all the rest in middle-income countries, with the top destinations being the US, Germany, and Saudi Arabia (IOM, 2019a). More than one tenth of the OECD population is foreign-born, which is approximately three times the global average. While it is true that most migrants are drawn to countries with higher income levels, this does not imply that migration necessarily occurs along the largest income gradients (Czaika and de Haas, 2012). One reason is that middle-income rather than low-income countries have the highest outflow rates, which is further discussed in the following subchapters. The most important sending countries are India, China, and Mexico (IOM, 2019a; OECD, 2020). As suggested by the regional trends, the distribution of foreign-born people will further diverge in future years and decades (IOM, 2019a).

The flows of migration did not always follow the current pattern, which can best be illustrated with the example of Europe. Formerly known as the continent of stagnant economies from where large flows emigrated to the Americas or Australasia in the 19th and early 20th century. In 1960 still, the share of Europeans in the overall migrant population exceeded three quarters. This changed in the second half of the 20th century when European economies developed high demands for labor. Today Europe is one of the most important migrant-receiving regions, whereas Europeans accounted for mere 22 percent of all international migrants in 2017 (Massey et al., 1993; Inglis, Li and Khadria, 2019; de Haas, Castles and Miller, 2020, p. 10). Another important development in the past 50 years took place in the Middle East and North Africa, where opportunities in the oil industry attracted massive inflows of workers. A similar development was caused by the growth of the labor-intensive manufacturing industry in many Asian countries (Inglis et al., 2019).

Overall, the past decades led to an expansion of migration flows from the global South to the global North even though South-South movements have remained significant, too (Koser, 2016; Ortega and Peri, 2013). Due to the increase in intercontinental movements the migrant population has gained in diversity. As a result, the visual and cultural differences between migrants and the host societies have become more visible, adding to the political salience of the topic (de Haas, 2007a; Massey et al., 1993). Changes have also been observed in the motivation for cross-border mobility with the rise of migrating students or pensioners who seek a better lifestyle or climate after their retirement (Inglis et al., 2019).

Even though the OECD (2020) has monitored a large drop in migration flows due to the COVID-19 pandemic, there are good reasons to assume that migration volumes will rebound rapidly and further rise in the near future. The first reason for this prediction is the ongoing population growth that is particularly pronounced in sending regions. The population of the African continent alone is expected to quadruple from 1 billion to 4 billion between 2015 and 2100 (Castelli, 2018; Clemens and Postel, 2018). At the same time, globalization and technological progress drive down migration costs and foster international networks in business and academia (Biffl, 2019; Docquier and Marfouk, 2006). Social media and the revolution of communication technology in general play a role by spreading (sometimes exaggerated) pictures of better living conditions in other parts of the world and facilitating the exchanges between migrants and their home communities (Biffl, 2019; Castelli, 2018). Besides, growing diasporas act as important magnets for populations in origin countries, representing self-perpetuating cycles where increases in diaspora size translate into the amplification of the pull effect, which in turn boosts the diaspora size (Castelli, 2018; Collier, 2013; Docquier and Marfouk, 2006). A more recently identified driver of potentially substantial cross-border movements is climate change whose importance is likely to grow in future years (Biffl, 2019; Castelli, 2018). However, this aspect along with concerns about conflicts and oppressive regimes falls under the humanitarian migration category that is not further discussed here. Lastly, I should point to economic development and income disparities as important factors, on whose role I elaborate in more detail in chapter 2.5. Besides the questions of how many people are on the move, it is also important to investigate who these people are and for which reasons they leave their countries, which is the subject of the following section.

2.3 Composition of migrants

While a lot of the attention in public discussions on migration is dedicated to international refugees, whose numbers have peaked to an unprecedented 26 million in 2018 according to the IOM (2019a), it is important to emphasize that two thirds of all international migrants are labor migrants, where the gender bias is particularly pronounced (58 percent are males vs. 52 percent in the overall migrant population). However, the current trends show an increasingly balanced gender representation (OECD, 2017; de Haas, Castles and Miller, 2020, p. 11). Moreover, migrants tend to come from urban rather than rural places, since life and work in cities help expand human capital and interpersonal networks. (Castelli, 2018; Clemens, 2020). An OECD (2017) study examining migration patterns in ten countries[6] with high emigration or immigration rates finds that many migrants are in their most productive years (mainly between 25 and 36) when leaving their country of origin. Emigrants were also found to have a higher likelihood of being employed before leaving and to be more skilled than the population average, which is also the state of knowledge in the literature (Castelli, 2018; Clemens, 2014; de Haas, 2007a). Nonetheless, the investigated immigrants were on average less well educated than the native-born population in the host society, which can be explained by different educational levels between sending and receiving countries.

[6] Households from Armenia, Burkina Faso, Cambodia, Costa Rica, Côte d'Ivoire, the Dominican Republic, Georgia, Haiti, Morocco, and the Philippines participated in the study.

As a consequence, migrants work more frequently in less desirable job positions (Biffl, 2019; Czaika and de Haas, 2012; OECD, 2017).

Regarding the skill levels of migrants, significant variations can be detected across destination regions. For example, North American and Australasian countries try to attract particularly high-skilled migrants, whereas immigration to Europe is much more centered around questions of asylum and family reunion. However, also European countries are increasingly aiming at recruiting skilled labor that is needed to foster the work force in many of Europe's aging societies (Docquier and Marfouk, 2006).

The brief overview of migration volumes, directions, trends, and the composition of migrants is needed to better understand the discussion on the relationship between development and migration as well as the discussion on opportunities and risks of people's movements which I summarize in the subsequent paragraphs.

2.4 Opportunities and risks of migration

Since my research interest is about ways to influence migration flows, a basic understanding of the associated consequences is required. Shaping an awareness of how migration – depending on its forms and volumes – may benefit or harm the receiving or sending countries (or the migrants themselves) helps understand the desire to steer migration and develop more appropriate policy recommendations. This subchapter is structured due to the different perspectives that need to be considered (global perspective, destination countries' perspective, sending countries' perspective, and the migrants' perspective).

2.4.1 Global perspective

To start with the global perspective, it appears surprising for how long the efficiency aspect of migration was neglected in the economic literature, considering that the origin of the economics discipline was tightly associated with studying the inefficiencies of trade barriers (Borjas, 2015; Clemens and Pritchett, 2019).[7] Hamilton and Whalley (1984) were the first ones to compute an estimate for the costs of the inefficient international allocation of labor. Their estimation that global production could be augmented drastically (their estimates ranged between increases of 50 to and 150 percent annually) through a higher degree of labor mobility were corroborated in a number of more recent studies (Benhabib and Jovanovic, 2012; Kennan, 2012; Klein and Ventura, 2009). These big hopes are questioned for example by Collier (2013), who fears that large-scale migration from poorer to richer countries reduces total factor productivity (TFP) in the rich world. He bases this argument on the assumption that migrants often come from societies with dysfunctional social models[8] and transfer economically harmful elements of these models through migration to countries with well-functioning models. An illustrative example of such potential spillovers is described in a study by Fisman and Miguel (2007). They found that the corruption level in the home countries of diplomats in New York (who enjoyed immunity until 2002) was highly correlated with the number of unpaid parking violations. Moreover, the authors uncovered that this behavior spread in the diplomat community with an increasing number of diplomats from low-

[7] Instead, most research efforts used to be targeted on the distributional consequences in destination countries (Clemens and Pritchett, 2019).

[8] Collier (2013, p. 32) adopts a very broad understanding of "social model" that encompasses institutions, narratives, norms, and organization.

corruption countries developing higher propensities for similar violations over time. Skepticism towards the size of the potential gains of liberalizing migration regimes due to harmful spillovers are shared by Borjas (2015). Related arguments by different authors about potential welfare decreasing aspects of migration are picked up when discussing the consequences for host societies.

Clemens and Pritchett (2019) criticize that Collier (2013) does not provide a clear theoretical model of how this transmission (the extent to which the arrival of foreigners drives down TFP) plays out. To my knowledge, they are the only ones who tried to quantify said spillovers. While, they do not make a case for entirely open borders, their estimates suggest that levels at which the negative marginal effects of a decline in TFP may offset the positive effects of more efficient labor allocation "would be far in excess of current rates". Some of the aspects raised in this first section on the global perspective are further explored when illuminating the situations in the receiving and sending countries in the following paragraphs.

2.4.2 Receiving countries' perspective

In reality, labor flows are not shaped by some global central planner but rather by national governments that usually have more influence on who enters than on who leaves (Koser, 2016). Thus, I continue with the receiving end of migration, i. e., with the impacts of immigration. First, receiving countries often benefit from immigrants economically because – as stated above – many migrants are in their most productive years when they arrive and on average better educated than their co-citizens at origin. Particularly high-skilled migrants boost innovation through the cross-fertilization of ideas. They also show a disproportionate propensity to found start-ups and successful companies (Biffl, 2019; IOM, 2019a). Others fill gaps in important sectors or do jobs whose execution tends to be undesirable for members of the host societies.[9] The large majority of migrants is employed and the labor market trend prior to the pandemic has shown further improvements (Biffl, 2019; IOM, 2019a; OECD, 2020).

Nevertheless, migrants' labor market outcomes differ considerably across countries with unemployment rates being much smaller in the talent-seeking countries Canada, the US or Australia than in Europe, where the unemployment rate among foreign-born people has remained 4 percentage points above the average of the native-born population (Docquier and Marfouk, 2006; OECD, 2020). The overall impact on wages through immigration seems rather negligible at least for moderate inflows of people (Collier, 2013; IOM, 2019a). However, even though many immigrants are more skilled than their compatriots at origin, they tend to compete mostly with other locals in low-skilled sectors, especially with the already existing immigrant population. A competition between these groups may also exist in the housing market (Biffl, 2019; Collier, 2013).

At the same time, important non-economic benefits are created by migrants. These are particularly visible in the areas of food, music, culture, and arts where immigration has produced unprecedented levels of goods diversity. Immigrants also contribute to host societies through political engagement on various levels or volunteer work and can facilitate the social and economic integration of newly arriving immigrants (IOM, 2019a). However, there is also a flipside to diversity, that can be linked to the efficiency concerns expressed above. Several studies suggest

[9] This type of jobs, that large parts of the native populations refuse to work in, is frequently described as "dangerous, dirty and demeaning" (Czaika and de Haas, 2012).

that high levels of diversity bring about negative consequences for social trust, the provision of public goods or the acceptance of redistributive institutions, which has negative consequences both for the native population but also for the immigrants themselves (Alesina and La Ferrara, 2000; Algan et al., 2011; Luttmer, 2001; Miguel and Gugerty, 2005). Moreover, immigrants are sometimes portrayed as security concerns, which is an old phenomenon that gained momentum after 9/11 and triggered the securitization of migration policies.[10] These concerns can go beyond questions of physical security and are often shaped and instrumentalized by policymakers (Koser, 2016; de Haas, Castles and Miller, 2020, p. 232 f.).

2.4.3 Sending countries' perspective

The migration debate is often centered around the needs and interests of the developed world, although the consequences of migration for origin countries deserves at least as much attention (Biffl, 2019). In this context, the probably most prevalent issue is the "so-called" brain drain, which is the flow of talented and often expensively trained people from developing to industrialized countries (Brock and Blake, 2015). This phenomenon has expanded very substantially since the second half of the 20th century with the outlet of the brain drain being mainly North America and Australia (Biffl, 2019; Lucas, 2019, p. 4). For a long time, the associated consequences were assumed to be clearly harmful for the societies left behind, as skilled labor moves from human capital scarce societies to countries that are already abundant in skilled labor (Docquier and Rapoport, 2012). Today, the assessment of brain drain is much more mixed though. The outflow of qualified labor particularly in the health and education sector can indeed complicate the provision of essential public services to their populations (Brock and Blake, 2015; Koser, 2016). At the same time, the prospect of migration opportunities motivates people to invest more in their human capital. This "brain gain" hypothesis was empirically tested for instance by Batista, Lacuesta and Vicente (2010) in Cape Verde – a country very strongly affected by the brain drain phenomenon – who found substantial positive effects on educational outcomes. Because many of the students who study harder with the objective of leaving in the future, end up staying in their home country, there is a positive effect on human capital at origin. Unfortunately though, when governments observe the exodus of large numbers of those in whose education they had previously invested, they may henceforth be more reluctant towards education investments (Collier, 2013; Docquier and Marfouk, 2006).

This seems to happen even though origin countries benefit from remittances which are defined by the (IOM, 2019a) as "financial or in-kind transfers made by migrants directly to families or communities in their countries of origin", and which play a more important role than ever in sustaining families and communities in origin countries. The magnitude of remittances rose from USD 126 billion in 2000 to USD 689 billion in 2018 which makes them the most important economic benefit of emigration for the origin country, largely outsizing official development assistance (IOM, 2019a). Households that receive remittances show higher propensities to invest in businesses and human capital (de Haas, 2007a; OECD, 2017). The prospect of remittance payments has even incentivized some countries to specialize in "exporting " skilled labor, the best example being the Philippines (de Haas, 2007a). Aside from remittances, emigrants also invest directly in their home countries, foster trade relations and transfer knowledge and information,

[10] Securitization of migration policies refers to a process of social construction in which migration is tightly linked to security concerns (de Haas et al., 2020, p. 10).

particularly – but not exclusively – when returning home (de Haas, 2007a; Docquier and Marfouk, 2006). In conclusion, the exodus even of skilled individuals can be positive if there is no shortage of these skills at origin. Brain drain seems especially problematic when it refers to profession groups that are urgently needed to provide essential services, for example in the health or education sector. The net effect of brain drain is more likely to be negative if emigration rates are very high (Collier, 2013; Schiff, 2006).

Besides economic considerations, emigration can support social and political progress in the sending countries, for example by boosting the role of women at home. Depending on the region, there can still be a very substantial gender bias in the emigrant population, and therefore women often become household heads and manage the incoming remittances, which provides them with larger degrees of autonomy, as the mentioned OECD (2017) study shows. Another study from Mali by Chauvet and Mercier (2014) finds that return migrants show higher voting participation and apparently spread their political norms also in their neighborhoods – including among co-citizens with low formal education. This has been interpreted as a positive impact of return migration on political norms at origin. Similarly, Batista and Vicente (2011) reveal that international migration increases the demand for adequate political accountability. The effects were particularly strong among return migrants who spent time abroad in states with good governance systems. Spilimbergo (2009) finds robust evidence that individuals who studied in a foreign democratic country are more likely to promote democratization at home. This is highly relevant given that large parts of the elites in developing countries have received at least some foreign education. Aside from return migration, Karadja and Prawitz (2018) claim that the existence of emigration as an "outside option" augments the bargaining power of citizens and workers vis-à-vis the elites which also impacts democratization processes and unionization rates positively. Based on these insights, migration may be particularly beneficial if emigrants return at some point, such that they can contribute to the political and economic progress of the country with the skills, ideas and values gained abroad.

2.4.4 Perspective of migrants and their families

The last perspective to consider is the one of migrants themselves and their families. When moving to higher income countries and finding employment there, migrants often earn a multiple of former wages. Migration comes however at the cost of being separated from one's friends and families as well as being exposed to discrimination, and often relative deprivation compared to the host society. Depending on the cultural distance between origin and destination countries, a lacking knowledge of the local costumes and language can also translate into a feeling of alienation (Helliwell et al., 2018). Nevertheless, the World Happiness Report of 2018 finds that, by and large, migrants are on average happier after migration. The positive association prevails when comparing migrants to their statistically appropriate counterparts at origin. This seems to be driven by the direction of migration flows which mostly run from poorer and less happy to relatively rich countries with happier populations. Even though the average migrant is not in general less happy than their co-citizens in the host society, a significant difference remains for those who move to the happiest countries. Moreover, the effects on life satisfaction of the families left behind are reported to be on average positive (Helliwell et al., 2018).

Of course, much more could be said about the advantages and problems of migration. At least for moderate levels of migration, most researchers agree that the benefits (significantly) outweigh the costs (Clemens and Pritchett, 2019; Collier, 2013; Koser, 2016). But in some cases increasing mobility may also be harmful, and thus policymakers should not debate whether migration is a good or a bad thing in general, but rather try to identify desirable levels of immigration and emigration and ways to design policies to harness the potential of migration optimally (OECD, 2017). It seems plausible that migration benefits are subject to declining marginal effects (e. g., remittances, innovation, diversity), whereas the marginal costs of migration might be increasing (e. g., the erosion of social trust and cooperation or the brain drain consequences). Even though (modest) migration increases on a global scale seem likely to bring about more positive than negative impacts, the situation can be different for specific sending or receiving countries (Clemens and Pritchett, 2019; Collier, 2013).

After the analysis of the trends as well as the risks and opportunities of migration, the desire to steer migration flows can be better understood and assessed. However, it has also become clear that steering should not be equated with reducing flows, as many countries may well benefit from an increase in people's mobility. Whatever the preferred quantities are though, successful steering undoubtedly requires a sound understanding of the matter, which I intend to establish in the following subchapter.

2.5 Relationship between development and emigration

Many statements in debates on migration, some of which made by senior-level policymakers or institutions uncover a lacking fundamental understanding of the relationship between migration and development. This is often illustrated in the widespread "addressing the root causes of migration"-narrative[11] (Clemens and Postel, 2018; Lucas, 2019). By "root causes", they usually refer to the socioeconomic situations in the countries of origin and economic advancement as a means to reduce migration numbers (Clemens, 2014; de Haas, 2007b). This theory is closely associated with the standard neoclassicist view that wage differentials between countries and how they relate to migration costs are the driving force behind migration. The idea that workers from labor abundant and capital scarce countries with low wages seek better opportunities for themselves and their families in more developed nations with higher capital endowments and wages has been one of the oldest and most influential theories in the international migration literature. According to this theory, migration represents a form of investment because costs (travel costs, fees for visas, etc.) are incurred to achieve long-term gains. The larger the gains, i. e., the income differentials, the higher are the returns to migration and accordingly the more people move. The theory also predicts a migration-driven convergence of wage levels across countries up to a global equilibrium wage, as the labor-capital ratios across countries should become increasingly similar (Clemens, 2014; Dao et al., 2018; Massey et al., 1993).

Indeed, this argument cannot be entirely dismissed and can help explain a certain segment of the development-emigration relationship (for higher-income countries). The view of migration as an investment continues to be shared by leading researchers today (Biffl, 2019; Collier, 2013; Czaika and de Haas, 2012). However, it is a somewhat narrow-minded view as it disregards highly influential aspects that are essential for understanding the evolution of emigration rates particularly for lower levels of development, which I discuss below. While there is little doubt that differences

[11] Clemens and Postel (2018) list a number of quotations underlining how widespread this idea is in policy circles. They also highlight the significant role of migration deterrence for donor institutions in Europe and the US.

in wages across countries matter, migration does not take place along the largest income gradients. Therefore, the empirical evidence on emigration flows (as well as the absence of income convergence) is not in line with the theory's predictions of a continuously negative relationship between income and economic growth (Czaika and de Haas, 2012; Ortega and Peri, 2013). Rather, the data suggests that the relationship between economic development and emigration rates follows the shape of an inverted U, with middle income countries showing the highest outflow rates. This "mobility transition curve" was initially proposed in Zelinsky's (1971) seminal paper, and has since been called migration transition (Gould, 1979), migration hump (Martin, 1993) or emigration lifecycle (Hatton and Williamson, 1994).

In a recent study, Clemens (2020) could present robust evidence that the described relationship holds not only in cross-country analyses, but is indeed the path which countries tend to follow over longer time periods. According to him, emigration rates increase up to income levels of approximately $10,000 PPP[12] per capita before the trend is reversed, but the increase is slowed down between $5,000 PPP and $10,000 PPP. The detected magnitudes are very substantial with the estimated emigration rate at $10,000 PPP being 2.5 times as high as at $1,000 PPP. Dao et al. (2018) find similar results in another recent study and emphasize that roughly two thirds of the world population live in countries for which economic growth should be associated with *higher* rather than lower emigration rates.

In my dataset, which I introduce in chapter 4, this pattern is also visible, as figure A-7 demonstrates. In the face of the empirical evidence, socioeconomic improvements in developing countries – as desirable as they are – should not be expected to lower migration rates. If anything, they are likely to spur emigration. The deviation of the empirical observations from the standard neoclassical predictions represents a puzzle, for which some explanations have been suggested.

The probably most common approach to make sense of the empirics is based on the notion of credit constraints. Given the considerable costs linked to migration[13] combined with the lack of access to credit markets in many developing countries, this aspect suggests a plausible explanation for why emigration rates in the poorest countries may be lower than for middle income countries. Indeed, it also helps understand why within a given country, the poorest population groups tend to be underrepresented among emigrants (Clemens, 2014; OECD, 2017; Dao et al., 2018).

A second aspect with particular significance for my research question are developments in the area of education that typically coexist with sustained economic growth. People with higher levels of education find better employment opportunities abroad and can more easily meet the criteria for work visas. The accreditation of diplomas and degrees is also mostly easier for middle-income rather than low-income countries (Clemens, 2020, 2014). Another close companion of economic development is a demographic transition. The decline in child mortality caused by greater wealth typically precedes lower fertility rates leading to augmented cohorts of young people who cannot be fully absorbed by domestic labor markets with often rigid wages. The rise in population density also fuels housing prices and can bring along congestion issues (Clemens, 2014; Hatton and

[12] "Power purchasing parity (PPP) is a money conversion rate used to express the purchasing powers of different currencies in common units" (National Institute of Statistics and Economic Studies (INSEE, 2021).

[13] Even though mobility has been largely facilitated throughout the past decades, migration costs can still be high especially in low-income countries. In some cases, costs of migration (rather than mobility in general) have also been increased through political decisions (Ortega and Peri, 2013).

Williamson, 2005). These arguments offer an additional explanation both for the upward trend of the first part of the mobility transition curve as well as for the age structure and the educational attainment of emigrants.

Another relevant structural change is the transition from agricultural production in rural places to manufacturing in urban areas. When farmers have to leave their homes anyways, the costs to move across the border are sometimes hardly higher than moving to the city (Clemens, 2014). Further, as mentioned above, the skillset and networks people develop when working in cities are helpful for migration purposes (Castelli, 2018).

Lastly, the very significant effect of existing diasporas must be stressed. They themselves may not explain what initiates the growing numbers of emigrants in poor countries, but they largely fuel the process once it is set in motion. This strong pull effect is rooted in the variety of ways through which diasporas facilitate the migration process (provision of information, support with social and economic integration, initial accommodation, etc.). For this reason, the migration costs of first-movers are significantly higher and might be effective in deterring potential emigrants. However, once the first emigrants have settled abroad, many more are likely to follow (Collier, 2013; Czaika and de Haas, 2012; Hatton and Williamson, 2005; Massey et al., 1993). Combining the evidence on the high proportion of the world population living in countries with income levels below the estimated emigration peak, the accelerated economic growth in many of the world's poorest regions as well as the growing diasporas, provides even more reasons to predict that migration numbers will continue to rise – not only in absolute, but also in relative terms (Castelli, 2018; Dao et al., 2018).

To conclude, a variety of reasons have been listed to explain the – on first sight – puzzling shape of the migration hump curve. However, while the overall pattern appears to be quite robust, there is a lot of variation in the data, and, as pointed out in the introduction, disentangling the effects of the various components of development has proven difficult (Clemens, 2020, 2014; Dao et al., 2018; Ortega and Peri, 2013). Thus, while it is not justified to generally propose economic growth to bring down migration numbers, the evidence on the mobility-transition curve does not necessarily imply that all investments in poorer countries' economic advancements influence emigration rates positively (Clist and Restelli, 2020). To promote a more differentiated perspective, I proceed by investigating the role of infrastructures in migration processes and start by summarizing some findings in the literature.

2.6 Literature review on infrastructure and migration

Infrastructures play a very significant role in countries' development processes, and have been defined for instance as "those basic facilities, structures and services that serve as a back bone for the development and economic wellbeing" (Kalu et al., 2014, p. 14) of societies. Nevertheless, their role has been largely neglected in the international migration literature (Albouy, Cho and Shappo, 2019). Besides their significance for economic development, infrastructures also promote social aspects, and thus a differentiation into economic ("hard") and social ("soft") infrastructures is sometimes made. Typically, transport, electricity, and telecommunications infrastructures are viewed as classic examples of the former, whereas education, health or cultural infrastructures are labeled as "social" (Fourie, 2006).

Both types have been subject to investigations with respect to their impact on migratory movements, which I describe in this section. The focus in the international migration literature has been set on education infrastructure, aspects of which were discussed in chapters 2.3 and 2.4. To briefly recap, education is widely recognized as a driver of migration, which can be due to a variety of reasons. For instance, better educated people are more likely to have the necessary resources (financial resources, information, networks) to migrate, they have a better chance to apply successfully for work visas and more positive employment prospects in destination countries (de Haas, 2007b; OECD, 2017; IOM, 2019a; Clemens, 2020). The brain gain proponents stress that the effects can go in both directions with the prospect of emigration serving as an important motivation for human capital investments. Typically, the role of education is discussed from a within-country perspective, whereas my research interest is to compare the effects of education systems across countries (Fargues, 2019; Feliciano, 2005; Schewel and Fransen, 2018).[14]

Apart from education, the existing literature does not provide a rich picture on the impact of infrastructures on *international* migration. Rather, the analysis is centered around within-country location choices. For example, Lall, Timmins and Yu (2009) investigate pull and push factors for movements from lagging to leading regions in Brazil. Besides, the better labor market opportunities in the well-performing regions (pull factor) they also identify the lack of service provision through inadequate infrastructures in the struggling areas, which often pushes people from more rural places to the cities. Kalu et al. (2014) focus on rural-urban migration in Nigeria and express their concern that the large imbalances in the provision of infrastructures between rural and urban regions promote an unprecedented as well as unsustainable city growth in the developing world. They propose the enhancement of rural infrastructures as an effective tool to slow down the speed of rural-urban migration. In both papers, the role of infrastructures is essentially limited to the provision of public services (i. e., amenity value), whereas the productivity aspects of infrastructures are not considered.

Fried and Lagakos (2017) select a different approach in their evaluation of an electrification program in rural Ethiopia. While they also find that improved electricity infrastructures in rural areas helps mitigate rural-urban migration flows, they regard higher agricultural productivity as the key channel that incentivizes residents to stay in their villages or even attracts people from other places. However, they also touch on the role of school, health, water, and road infrastructure. The work of Shilpi, Sangraula and Li (2014) stands out as it clearly distinguishes the amenity effects of infrastructures from the effect on income using a two-stage estimation procedure. They investigate the sorting of migrants to locations in Nepal and find statistically significant preferences for access to paved roads and electricity.

These examples highlight that both an economic and a social/amenity dimension are entailed *within* most kinds of infrastructure. While life outside the workplace is evidently facilitated through good electricity and transport systems, education and health services shape human capital and thus people's productivity (Fourie, 2006). Therefore, the above-mentioned distinction between social and economic infrastructures is little convincing. In the following chapter I develop a novel classification from which clear consequences for people's incentives to migrate can be derived.

[14] In the within-country approach, the question of what kind of people of a society emigrate (e. g., the ones with above-average educational attainments compared to their compatriots) typically is dominant.

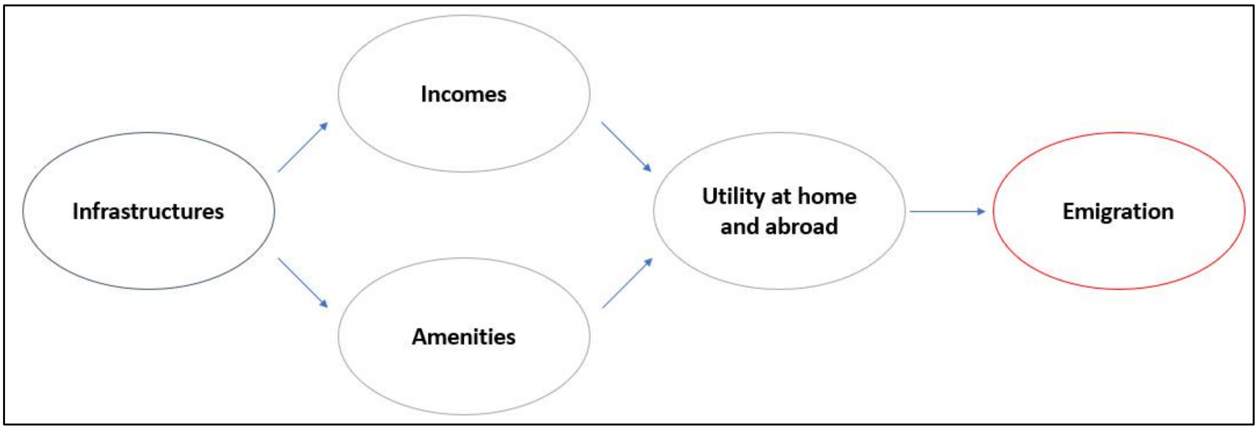
3 Theoretical framework of the infrastructure-emigration relationship

I lay out my proposal for a novel classification of infrastructures, that is useful to analyze migration movements, in two stages. First, I introduce the core ideas and try to illustrate them with economic intuitions. The second part translates these considerations into a formal model, from which the concrete hypotheses are derived.

3.1 New classification of infrastructures

The literature overview presented in chapter 2.6 highlighted some evidence on the role of infrastructures for people's location choices as well as the two important channels through which they unfold their impact: productivity and amenity values. Shilpi, Sangraula and Li (2014) define the former as a shift in the production possibility frontier, i. e. the production capacities given a fixed endowment of production factors (e. g., electricity supply helps with automation processes). On the other hand, infrastructures also enable "households to carry out essential chores efficiently and to enjoy leisure more fully" (Shilpi, Sangraula and Li, 2014, p. 3), which is understood as the amenity value. While I do not depart from the view of migration as an investment decision, I follow the recommendation of Czaika and de Haas (2012) to broaden the perspective and look beyond income disparities to generate a better understanding of why people move through adopting amenity values in my framework. Figure 2 summarizes how infrastructures influence utility through the mentioned channels. The exact connection to utility levels at home and abroad as well as the resulting consequences for emigration are illustrated in the following paragraphs.

Figure 2: Channels from infrastructures to utility and emigration

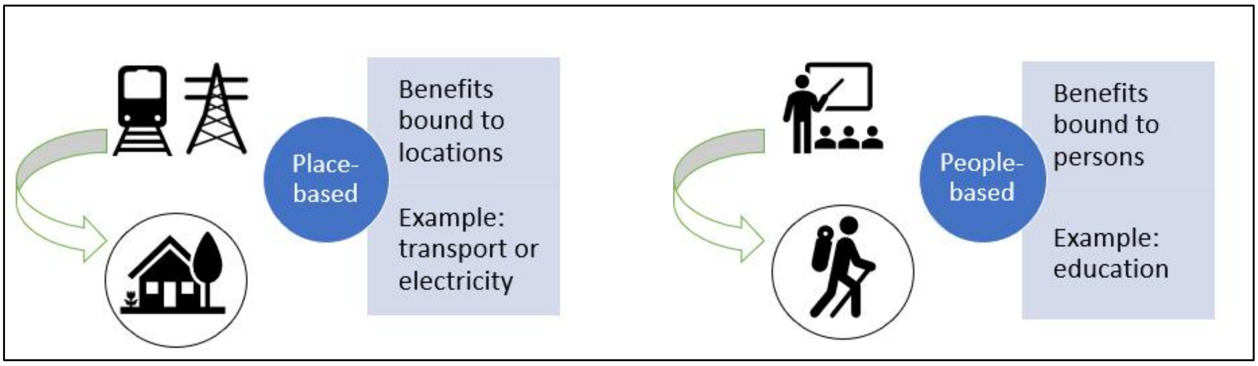


Source: Own compilation.

Even though the grouping of infrastructures according to my conceptional framework is somewhat similar to the groups resulting from the distinction between economic and social infrastructures, the underlying criterion is entirely different. Inspiration comes from the work of Adhikari and Gentilini (2018) who distinguish between place- and people-based designs of social assistance programs and evaluate their effects on emigration in a meta study. Transferring this distinction to the field of infrastructures, I differentiate between place-based infrastructures (PLI) and people-based infrastructures (PEI). While both infrastructure types are income- and amenity-enhancing, they differ in how they do so. The provision of PLI (e. g., electricity or transport infrastructures) boosts productivity and raises quality of life for – and only for – the people living in locations connected to these infrastructures, i. e., the electricity grid or the road/railroad network. However, once a person decides to leave this location, they no longer benefit from the past delivery of services.

This is different for PEI (e. g., education), where the service *delivery* may also be bound to specific places (school catchment areas), but these services transform into "human capital" from which the person can benefit in their future regardless of where they live or work. As Dustmann et al. (2011) put it, "[h]uman capital cannot be separated from its owner" and thus it drives up personal utility through incomes and higher quality of life outside the workplace either in the country of origin or abroad. The human capital concept here is extremely broad, as it also captures utility-enhancing aspects that are independent of a person's productivity (e. g., musical education for hobby musicians). The key difference in the characteristics between the two types are depicted in figure 3.

Figure 3: Characteristics of PLI and PEI



Source: Own compilation.

Even from the narrow perspective of the income-centered neoclassical theory, this distinction reveals why a definition of income differentials based on the differences of average incomes by country is too short-sighted. The development of PLI in origin countries affects exclusively incomes and amenities at origin, whereas the prospects of potential emigrants for incomes and amenities abroad remains unaffected. Therefore, the differentials vis-à-vis possible destination countries decline (under the ceteris paribus condition). Conversely though, a better PEI quality does not only spur incomes and amenities at home but also in potential destination countries. If the returns to education are higher abroad than at home, the income differentials from the perspective of an individual may even widen while at the same time the average wages by country converge.[15] This is because the change in *potential* incomes in the case of emigration remains unobserved.[16]

Figure 4 illustrates these mechanisms graphically. On the left-hand side, the scenario of PLI development is displayed. The blue bar that represents the individual's income in the case of staying goes up when PLI improves, whereas the income for leaving (orange bar) remains unaffected. Accordingly, the differentials decline. Conversely, in the case of PEI enhancement, incomes are positively affected regardless of the place of residence. Therefore, the change in the

[15] There is evidence that more educated people are not just more likely to migrate but also tend to settle in destinations with relatively high returns to skill (Grogger and Hanson, 2008).

[16] The positive effects of an increased productivity of emigrants on the average incomes in the destination country are considered negligible.

differentials depends on the returns to education. In the following subchapter, I argue why these returns are likely to be lower at home than abroad.

Figure 4: Effects of PLI and PEI on income differentials

Source: Own compilation.

This classification is helpful for the analysis of the development-migration association for the following reasons. First, it creates a clear link between infrastructures and emigration decisions and can contribute to explaining the positive slope in the first part of the mobility-transition curve: In view of the tight association between economic and human capital growth, this framework highlights that converging aggregate income levels, do not necessarily imply declining differentials from the perspective of the individual (Clemens, 2020). Moreover, it offers a more comprehensive approach by also considering the amenity value of infrastructures. Differences in PLI and PEI across countries might add insights into the very substantial variations around the mobility-transition curve, too.

The public policy relevance of the infrastructure aspect stems from the scope for political influence in this field. Since the necessity for poor countries to raise income levels is undisputed and many of the migration-driving structural changes (e. g., demographics, urbanization, or the growth of diasporas) are so tightly linked with economic development, the room for policy influence is very limited (Clemens, 2020). Of course, infrastructure development is also closely associated with economic growth, but the decisions on the design of infrastructure development typically fall under the public domain and thus offer some leeway for governments to exert a certain degree of influence on migration processes.

It is important to note that infrastructures can also affect the cost side of migration. This may refer to both the actual act of moving (i. e., information and travel costs) as well as costs associated with living in a foreign society, far away from one's family and friends. This aspect of alienation seems to be a larger obstacle to migration for many people than the costs of movement itself (Lucas, 2019). Unfortunately, the available data does not allow me to distinguish between the different effects on incomes, amenities, and migration costs. As it is my objective to investigate the implications associated with the infrastructure types, I try to minimize distortions due to changes in costs. Thus, I do not include ICT infrastructures which can plausibly make the separation from home communities a lot less painful. Moreover, I select transport infrastructures that are dominant in within-country mobility. Given the available options in the GCI dataset, from which I obtain my infrastructure data, I adopt road, railroad, and electricity infrastructures to analyze the impact of PLI.

PEI candidates are more difficult to find. Besides education, health infrastructure is a second candidate. Unfortunately, the health variables in the dataset are not related to health infrastructures but rather health outcomes, which depend on all kinds of factors, not just the respective infrastructure (e. g., nutrition, climate, regional prevalence of diseases, education, etc.).[17] Therefore, health infrastructure data would need to be collected from a different source, which would raise the question of which indicator to choose best (density of hospitals/pharmacies, health expenses or education of health workers, etc.) and cast doubt on the comparability with the other variables that are all measured identically (see chapter 4). For these reasons, a more in-depth analysis of the relevant health infrastructure features would be required before adopting the variable.

As a result, the infrastructures selected for PLI are road, railroad, and electricity, while PEI is represented merely by education. With the presented intuition and the selection of infrastructures in mind, I now introduce a model that formalizes the outlined channels and allows for a clear derivation of hypotheses.

3.2 Formalization of the infrastructure-emigration relationship

I select a discrete choice model, in which an individual optimizes their utility by deciding whether to emigrate to some destination country or to stay in the home country. As it is not the intention of the paper to predict which country a migrant goes to, the destination country can simply be regarded as the "most promising" foreign country in terms of the individual's utility. I make the assumption of a single and irreversible migration decision that is taken at a specific point in time, the consequences of which are discussed in chapter 6.1.

The individual's life is divided into three distinct time periods: the human capital formation period (*t-1*), the migration period (*t*), and the working period (*t+1*). In *t-1*, the potential emigrant receives education whose quality depends on the education infrastructure in the country of origin ($I_{O,t-1}^{PE}$). This period is followed by a period *t* in which the individual decides if they prefer to stay and work at origin or to leave and work abroad. The decision is based on comparing income and amenity prospects at origin and destination in the working period *t+1* with migration costs. Individuals are perfectly informed about future outcomes (incomes and amenities) as well as about migration costs. The condition for emigration is that the discounted benefits of leaving outweigh the costs of leaving which is expressed formally in the inequality:

$$\gamma * \left(u_{i,t+1}^{L} - u_{i,t+1}^{S}\right) > c_t. \qquad (1)$$

The utilities of spending the period *t+1* at home or abroad are represented in the terms $u_{i,t+1}^{S}$ and $u_{i,t+1}^{L}$, respectively, while c_t represents the costs of migration in *t*. γ is the discount factor that is assumed to be constant across individuals.

Both utilities at home and abroad are functions of infrastructure quality. The utility of staying depends positively on income ($y_{i,O,t+1}$) and amenities ($a_{i,O,t+1}$) at origin which are in turn dependent on PLI ($I_{O,t+1}^{PL}$) and PEI ($I_{O,t-1}^{PE}$) in the country of origin. Utility of leaving, on the other

[17] The health variables included are the number of Malaria and Tuberculosis cases per 100,000 population, HIV prevalence and the business impacts of these diseases. Moreover, measures for infant mortality and life expectancy are included (World Economic Forum, 2021).

hand, depends on income ($y_{i,D,t+1}$) and amenities ($a_{i,D,t+1}$) at destination which are functions of PLI at destination ($I_{D,t+1}^{PL}$) and of PEI at origin ($I_{O,t-1}^{PE}$), as outlined in equations (2) and (3):[18]

$$u_{i,t+1}^{S} = g\left(y_{i,O,t+1}\left(I_{O,t+1}^{PL}, I_{O,t-1}^{PE}\right), a_{i,O,t+1}\left(I_{O,t+1}^{PL}, I_{O,t-1}^{PE}\right)\right) \quad (2)$$

$$u_{i,t+1}^{L} = h\left(y_{i,D,t+1}\left(I_{D,t+1}^{PL}, I_{O,t-1}^{PE}\right), a_{i,D,t+1}\left(I_{D,t+1}^{PL}, I_{O,t-1}^{PE}\right)\right) \quad (3)$$

To make the mechanisms and the derivation of my hypotheses clearer, I develop a more specific version of the model. My objective is first to model how changes in infrastructures impact utilities through income and amenities. The assumption is made that utility increases monotonously in both. For the sake of simplicity, I follow Shilpi, Sangraula and Li (2014) or Grogger and Hanson (2008) by approximating utility with a linear function of income and amenities, even though linearity is not necessary for the derivation of my hypotheses.[19] Hence, I define the individual's utility in period *t+1* as:

$$u_{i,t+1} = \alpha * y_{i,t+1} + \beta * a_{i,t+1} + \epsilon_i \quad (4)$$

α and β are weights associated with incomes and amenities. I assume potential migrants to be identical in the sense that they do not differ in how they value income and amenities. Throughout the model, weights are assumed to be constant and to take values between 0 and 1. ε_i captures unobservable idiosyncratic chararacteristics of the individual.

Defining the relationship between wages/amenities and infrastructures, I draw on Clemens (2020) who models income as a function of a country-specific base wage and a term that rewards the individuals for years of schooling. Similarly, I assume incomes and amenities to depend on a base income/amenity value capturing all kinds of relevant country-specific characteristics as well as infrastructure-specific terms that raise these base values:

$$y_{i,t+1} = \overline{y_R} + \delta * \left(I_{R,t+1}^{PL}\right)^{\lambda} * \left(I_{O,t-1}^{PE}\right)^{1-\lambda} \quad (5)$$

$$a_{i,t+1} = \overline{a_R} + \eta_1 * I_{R,t+1}^{PL} + \eta_2 * I_{O,t-1}^{PE} \quad (6)$$

Subscript *R* stands for the individual's place of residence, be it the country of origin or the destination country. The key difference between equations (5) and (6) is that I assume a complementary nature of infrastructures in the wage context while their contributions to the amenity value are considered independent. This follows the rationale that even the best educated workers would be highly unproductive in the case of (largely) inexistent PLI, e. g., because a minimal provision of electricity is essential for most working processes or because workers need to commute to factories. Similarly, top quality endowments regarding roads, railroads or electricity lines would be of very little (productive) value if workers lacked even most basic skills.[20] Conversely, the amenity equation assumes that the amenity contribution of PLI does not

[18] Note the differences in the time index. While human capital is formed in t-1, PLI quality matters in t+1 when the individual is working.
[19] A monotonous increase is sufficient.
[20] The link to the economic literature can be made by considering production functions with Total Factor Productivity (TFP) and human capital (see Abbasa and Foreman-Peck (2007), for example). Considering PLI as a proxy for TFP and education infrastructure as a proxy for human capital, I follow the structure of a multiplicative combination of TFP and human capital.

systematically depend on the quality of the education system and vice versa.[21] The parameters δ, λ, η_1 and η_2 are weights for the infrastructures' impacts.

Migration costs capture all kinds of material and immaterial aspects related to the process of moving as well as living in a foreign society. They are modelled as a constant term rather than as a function of infrastructure, which represents another simplification. As argued above, I reflected this aspect in the selection of infrastructures, for instance by not adopting ICT infrastructure. Accordingly, the infrastructures included affect rather the one-shot costs of moving than the long-term costs of living in a foreign society and I assume that the latter are small relative to the sustained benefits through higher income and amenity values. I want to stress that I do not claim total migration costs to be generally negligible (which they are not). Rather, I compare the *changes* associated with infrastructure improvements.[22] This assumption serves to derive an unambiguous hypothesis on the effects of PLI as I point out below. Combining my considerations yields the following utility functions:

$$u_{i,t+1}^{S} = \alpha * \left[\overline{y_O} + \delta * \left(I_{O,t+1}^{PL}\right)^{\lambda} * \left(I_{O,t-1}^{PE}\right)^{1-\lambda}\right] + \beta * \left[\overline{a_O} + \eta_1 * I_{O,t+1}^{PL} + \eta_2 * I_{O,t-1}^{PE}\right] + \epsilon_i \quad (7)$$

$$u_{i,t+1}^{L} = \alpha * \left[\overline{y_D} + \delta * \left(I_{D,t+1}^{PL}\right)^{\lambda} * \left(I_{O,t-1}^{PE}\right)^{1-\lambda}\right] + \beta * \left(\overline{a_D} + \eta_1 * I_{D,t+1}^{PL} + \eta_2 * I_{O,t-1}^{PE}\right) + \epsilon_i \quad (8)$$

These formulas provide insights into the relations between infrastructures and place-specific utility, which serves as the foundation for the decision to emigrate (see figure 2). Inferring precise hypotheses from the model, requires one further assumption: The quality of PLI in the receiving country is expected to be higher than in the sending country. In this context, I would like to refer to chapter 2.2, where I described that most migrants target countries with higher income levels than at origin. Combining this evidence with the high correlation between income levels and infrastructure quality (see chapter 5.2), this assumption is likely to hold for the large majority of migratory movements.

The following derivatives matter for inferring my hypotheses:

$$\frac{\partial u_{i,t+1}^{S}}{\partial I_{O,t+1}^{PL}} = \alpha * \delta * \lambda * \left(I_{O,t+1}^{PL}\right)^{\lambda-1} * \left(I_{O,t-1}^{PE}\right)^{1-\lambda} + \beta * \eta_1 > 0 \quad (9)$$

$$\frac{\partial u_{i,t+1}^{L}}{\partial I_{O,t+1}^{PL}} = 0 \quad (10)$$

$$\Rightarrow \frac{\partial u_{i,t+1}^{S}}{\partial I_{O,t+1}^{PL}} > \frac{\partial u_{i,t+1}^{L}}{\partial I_{O,t+1}^{PL}}$$

[21] This is not to say that the benefits are of similar nature across countries. In states with splendid school systems, electricity may be used for very different consumption purposes than in countries with poor education systems. However, it is not evident why people in countries with better schools benefit to a larger extent from physical infrastructures than people in states with poor education systems. To my knowledge, models that describe the relationship between the quality of infrastructures and the amenity value in a country do not yet exist in the literature.

[22] Of course, an analysis of the effects on migration costs would also belong to a comprehensive analysis of the role of infrastructures but go beyond the scope of this paper.

PLI influences the utility of staying positively (all terms on the left are positive), while the utility of leaving remains unaffected. Since migration costs are assumed to remain unaffected, it can be concluded that better PLI should incentivize people to stay.

Hypothesis 1 (H1):
PLI has a negative effect on emigration.

The second hypothesis concerns the relation between PEI and emigration. Improvements of PEI drive up utility at origin as well as destination through higher wages and amenities:

$$\frac{\partial u_i^S}{\partial I_O^{PE}} = \alpha * \delta * (1-\lambda) * \left(I_{O,t+1}^{PL}\right)^{\lambda} * \left(I_{O,t-1}^{PE}\right)^{-\lambda} + \beta * \eta_2 > 0 \tag{11}$$

$$\frac{\partial u_i^L}{\partial I_O^{PE}} = \alpha * \delta * (1-\lambda) * \left(I_{D,t+1}^{PL}\right)^{\lambda} * \left(I_{O,t-1}^{PE}\right)^{-\lambda} + \beta * \eta_2 > 0 \tag{12}$$

It can be shown that the positive effects at destination outsize the ones at origin:

$$\alpha * \delta * (1-\lambda) * \left(I_{D,t+1}^{PL}\right)^{\lambda} * \left(I_{O,t-1}^{PE}\right)^{-\lambda} + \beta * \eta_2 > \alpha * \delta * (1-\lambda) * \left(I_{O,t+1}^{PL}\right)^{\lambda} * \left(I_{O,t-1}^{PE}\right)^{-\lambda} + \beta * \eta_2$$

$$\Rightarrow \left(I_{D,t+1}^{PL}\right)^{\lambda} > \left(I_{O,t+1}^{PL}\right)^{\lambda}$$

$$\Rightarrow \frac{\partial u_i^L}{\partial I_O^{PE}} > \frac{\partial u_i^S}{\partial I_O^{PE}}$$

Under the above assumption of $\left(I_{D,t+1}^{PL}\right) > \left(I_{O,t+1}^{PL}\right)$, the utility of leaving grows more than the utility of staying, which incentivizes emigration.

Hypothesis 2 (H2):
If PLI quality at destination is higher than at origin, PEI has a positive effect on emigration.

This hypothesis is driven by the assumed complementary character between the two infrastructure types that boosts the PEI effect on wages for higher levels of PLI. In light of the dominant direction of migration flows, I assume that PLI quality at destination is overall higher than at origin. Therefore, better education in sending countries is expected to widen the gap in utility prospects. Mathematically, this can be expressed by taking the derivative of the PEI effect on utility with respect to PLI:

$$\frac{\partial u_{i,R,t+1}}{\partial I_{O,t-1}^{PE}\,\partial I_{R,t+1}^{PL}} = \alpha * \delta * (1-\lambda) * \lambda * \left(I_{R,t+1}^{PL}\right)^{\lambda-1} * \left(I_{O,t-1}^{PE}\right)^{-\lambda} > 0 \tag{13}$$

Since all terms on the left-hand side are assumed to be positive, their product must be positive too. Confirming this interaction in the data does not constitute an independent hypothesis but would give support to the model structure and the assumed channels underlying the second hypothesis. Intuitively, this implies that relatively good education systems coupled with poor PLI quality are especially prone to stimulate emigration as workers are all the more prevented from realizing their potentials. Therefore, I will also test this "sub-hypothesis", as a driver for H2.

Table 2 summarizes the two main hypotheses that are tested in the following empirical section.

Table 2: Summary of hypotheses

H1	*PLI has a negative effect on emigration.*
H2	*If PLI quality at destination is higher than at origin, PEI has a positive effect on emigration.*

The formal derivation of clear hypotheses allows me to turn to the empirical part of the paper, in which the hypotheses are to be tested.

4 Empirical specification and data

This chapter serves to prepare the empirical investigation of the hypotheses. To this end, I derive an empirical model that is based on the theoretical framework as well as the literature and identifies the set of variables needed for the subsequent analyses. I continue by presenting my data.

4.1 Empirical model

The empirical model relates the outcome variable with the set of independent variables needed for the research design. In this case, the outcome variable is the net emigration rate by country and five-year period. Road, railroad, and electricity infrastructure (representatives of PLI) as well as education infrastructure (representative of PEI) constitute the central independent variables. Following the theoretical model regarding the distinction between different time periods, it would be desirable to work with lagged values for education. However, due to availability constraints of infrastructure data, the use of lagged values would mean losing all observations for the first of three periods. As the variations of the infrastructure scores per country are small over time (see chapter 5.4.1), I avoid this highly undesirable consequence by sticking to the education score of the current period. An interaction term for PLI and PEI (i. e., education) is included to test the mentioned sub-hypothesis.

Further, a variety of controls are adopted in the model. The first set of controls is meant to take account of forms of forced migration due to conflicts, a lack of freedom or democratic rights, homicides as a proxy for extreme physical violence or natural disasters (Clist and Restelli, 2020). The reason for their inclusion is that forced migration follows a logic distinct from the presented mechanisms. The second set of controls consists of median age, population density, urbanization, and the migrant stock at the beginning of each period, whose roles in the emigration context were discussed in chapter 2.5.

The last included control variable is gross national income (GNI) which represents a particular case. The reason is that part of the channel through which infrastructures are expected to influence emigration are higher earnings at home, and GNI can be viewed as a proxy for the income situation of households in the respective country. At the same time though, GNI represents the state of economic development of a country, which may capture multiple relevant third variables (e. g., the importance of credit constraints, telecommunication system, income disparities vis-à-vis potential destination countries, etc.). The inclusion of GNI still allows to test if two countries with similar national income levels differ in their emigration rates depending on the quality of their PLI and education infrastructure. Therefore, it is always included in the last model specifications, i. e., in specifications (4) and (5) (see chapter 5).

Another important aspect for the set-up of the empirical model is the data type which has consequences for the notation. Fortunately, data for the stated variables can be found in different years, which allows me to work with panel data. In my case, data for three time periods are available (see following subchapter). The use of panel data (also known as cross-sectional time series data) is increasingly popular among researchers in the field of economics and the social sciences, as it is associated with a number of advantages compared to pure cross-sectional designs. One advantage is the higher number of observations available for the analysis. Further, it facilitates the identification of causal relationships by adding before-and-after observations, which is one key component of causality. The stability of the relationship between outcome and independent variables can be observed over time and unobserved factors can be controlled for by including fixed effects (Xu, Lee and Eom, 2007, p. 573 f.).

Combining my decisions on variable selection and data type yields the following regression equation:

$$\begin{aligned} Emigration_{it} = {} & \beta_o + \beta_1 PLI_{it} + \beta_2 PEI_{it} + \beta_3 (PLI_{it} * PEI_{it}) + \beta_4 Conflict_{it} + \beta_5 Freedom_{it} \\ & + \beta_6 Homicides_{it} + \beta_7 NaturalDisasters_{it} + \beta_8 MedianAge_{it} \\ & + \beta_9 PopulationDensity_{it} + \beta_{10} Urbanization_{it} + \beta_{11} MigrantStock_{it} \\ & + \beta_{12} GNI_{it} + u_{it} \end{aligned}$$

The indices *i* and *t* serve to identify entity and observation period. This is the regression equation for a random effects model. If however fixed effects are introduced, additional parameters must be added that account for unobserved heterogeneity (Xu, Lee and Eom, 2008, p. 575). The corresponding regression equation is:

$$\begin{aligned} Emigration_{it} = {} & \beta_o + \beta_1 PLI_{it} + \beta_2 PEI_{it} + \beta_3 (PLI_{it} * PEI_{it}) + \beta_4 Conflict_{it} + \beta_5 Freedom_{it} \\ & + \beta_6 Homicides_{it} + \beta_7 NaturalDisasters_{it} + \beta_8 MedianAge_{it} \\ & + \beta_9 PopulationDensity_{it} + \beta_{10} Urbanizaiton_{it} + \beta_{11} MigrantStock_{it} \\ & + \beta_{12} GNI_{it} + \alpha_i + \lambda_t + v_{it} \end{aligned}$$

α_i and λ_t represent the fixed effects for countries and time periods. Their role and the exact differences between the random effects and the fixed effects model are discussed below. The empirical model serves as the foundation for the subsequent regression analyses. It also indicates what variables are required to test my hypotheses, which is why I continue in a next step with describing my data and the respective sources.

4.2 Data sources

Before speaking about the origin of the data though, I address the question of which observation units should be included. The objective of this piece of research is to collect evidence for or against the stated hypotheses by taking a very broad perspective rather than analyzing concrete case studies. Despite all kinds of differences between nations, there is no evident reason to exclude certain groups of countries ex ante, as the outlined mechanisms are at least in principal universal.

For this reason and to maximize the number of observations, I adopt all but the tiniest countries[23] in my analyses and focus on the period from 2005 to 2020. Table A-1 lists all 120 countries included in the analysis. The overall observation period is divided into the (sub-)periods 1 (2005 to 2010), 2 (2010 to 2015), and 3 (2015 to 2020).[24] The limitation to this time horizon is a consequence of data availability for infrastructures (see below).

For a long time, research on emigration was a challenging endeavor as data used to be scarce or of low quality (Clemens, 2020). However, recent improvements have substantially improved the data situation. Building up on an initiative of the University of Sussex regarding the creation of a global origin-destination database, the World Bank's (2021a) "Global Bilateral Migration Database" as well as UN DESA (2021a) publish comprehensive data on migrant stocks. As the UN data covers a more recent period (1990-2019) and updates the data every 5, rather than 10 years, I prefer this data source for my research purposes. The estimates for the migrant stocks are mainly obtained from population censuses supplemented with data from population registers or nationally representative surveys and refer to 1st of July of the respective year. Whenever the information is available, the definition of a migrant follows the introduced place of birth criterion. In the absence of this piece of information, the classification of migrants is based on citizenship. Overall, 232 countries or areas are included (UN DESA, 2021a).

Since I am interested in migration flows rather than migrant stocks, I derive the flow data from the foreign-born populations in the years 2005, 2010, 2015 and 2019. Following Clemens (2020) or Dao et al. (2018), I simply use the differences between the migrant stocks in successive periods to approximate net emigration flows. Net emigration is defined as the number of people born in country A who emigrate from A, minus the number of people born in A who return to A (after living abroad). These differences are then divided by the population size in the initial year of the respective five-year period. Because the most recent stock data is given for 2019 rather than 2020, this period is one year shorter than all previous ones. To account for that, I multiply emigration of this period by 1.25. Unfortunately, migrant stocks do not only change as a consequence of cross-border movements, which is discussed in chapter 6.2. This net emigration flow relative to population size constitutes the central outcome variable of my research design. The share of the migrant stock in the domestic population is used as a control variable approximating the diaspora size.

High quality infrastructure data are taken from the WEF's (2021) GCI. This index monitors 12 essential pillars for the competitiveness of 137 countries around the globe that account for 98 percent of global GDP. The GCI collects information from internationally recognized organizations (IMF, World Bank, UN agencies) complemented by data from the WEF's Executive Opinion Survey, which is also the source for my infrastructure variables. The survey has been set up to improve data availability in areas known for data scarcity or unreliability and draws on the

[23] With tiniest countries I refer to nations with a population of less than one million people. Besides the obvious reason of data scarcity, which makes the inclusion for most cases impossible anyways, countries with tiny populations often have oversized international migrant stocks that generate very particular migration patterns that may dominate other factors (Dao et al., 2018).

[24] On first sight, this period definition gives the impression of overlapping 6-year periods. The reason is though that the migrant stock data always refers to the migrant stock size at the middle of a given year. For all other variables, which follows annual structures, the averages of the years 2005 to 2009 (for period 1), 2010 to 2014 (for period 2), and 2015 to 2019 (for period 3) are selected.

responses from many thousand business leaders in a multitude of countries.[25] It is comprised of 150 questions, clustered in 15 sections, most of which ask respondents to give their evaluation on a specific topic (Schwab, 2017).

The second pillar of the index measures the quality of a variety of infrastructure variables, among them are “Quality of roads”, “Quality of railroad infrastructure” as well as “Quality of electricity supply”. The item “Quality of the education system” is part of pillar 5. All these variables are to be assessed on a scale ranging from 1 (lowest rating) to 7 (highest rating). The exact phrasing of the questions can be found in Table A-2.

As I argue in the following subchapter, I form an index for PLI that includes the scores for road, railroad, and electricity infrastructure by following these steps: First I take the averages of the five-year periods of each variable individually. Then, I standardize these averages (z-standardization) before adding them up and dividing the sum by three. This way, the index is coined similarly by all three of its components. Finally, I standardize the variable once more to facilitate the interpretation and the comparison with the PEI variable during the analyses of the results.[26]

The main drawback of this data source is the restriction to the period from 2007 to 2017. Data from more recent years exist but are not comparable due to changes in the data collection process (Schwab, 2018). I use the available data to create average values for the three five-year periods (2005-2009, 2010-2014, and 2015-2019), which roughly correspond to the migration flow periods. Evidently, the first and third period are not fully covered which is why the period averages are an approximation. This procedure seems acceptable due to low degrees of volatility over time (see chapter 5.4.1).

Control variables are included, for example, to account for forms of forced migration. To control for conflicts, I include a variable representing the conflict or terrorism deaths. The data comes from the “Global Burden and Disease Data Base” of the Institute for Health Metrics and Evaluation (2021) and is expressed as the number of persons killed through conflict or terrorism per 100,000 population. Another relevant aspect is the extent of liberty. A good indicator for the degree of freedom by country is published in the “Freedom in the World” reports by Freedom House (2021). The freedom scores are expressed in numbers from 1 (very free) to 7 (unfree). Note that higher scores in this variable mean lower levels of freedom. Forced migration may also be due to a generally high prevalence of violence (independent of large-scale conflicts) or natural disasters. The respective data (victims of homicides per 100,000 and number of people affected by natural disasters per 1,000 population, respectively) is obtained from the United Nations Office on Drugs and Crime (2021) and from the “Emergency Event Database” of the Centre for Research on the Epidemiology of Disasters (2021).

Further control variables that account for the median age, the population density and the degree of urbanization of the respective country are published by UN DESA (2018, 2019b). I choose Gross National Income (GNI) per capita in constant and international dollars as control for a

[25] The 2017-2018 report collected data from 12,775 businesses from 133 countries (Schwab, 2017).
[26] Since the education variable (averaged across the five-year periods) happens to have a standard deviation of 1, the comparison between the two becomes very intuitive.

country’s income level, as this measure controls for price level differences over time and across countries. The data comes from the World Bank's (2021b) World Development Indicators.

The instrumental variable regression requires two additional variables. The instrument itself, wetland, is taken from an OECD (2021) dataset on different types of land cover and indicates the share of wetland in a country’s territory. To control for alternative channels, as I discuss in chapter 5.3.4, I also adopt a variable measuring the value of agricultural production by country. [27] This data comes from the UN Food and Agriculture Organization (2021).

The empirical model combined with an overview of the dataset is the necessary equipment for the empirical investigations following in the next sections.

5 Methodology and results

This chapter introduces various statistical methods for an empirical investigation of the stated hypotheses. I explain the different tools, justify their applications, present and interpret the results. Most insightful are the multivariate regression analyses, whose general methodology is laid out in chapter 5.3.1. However, when interpreting the results, I also hint at drawbacks of the respective design and explain how a different approach may address the issue. Therefore, methodological elements are also scattered in the sections that discuss the findings. In a first step though, descriptive statistics and bivariate relationships are briefly outlined to create a first impression for the data structure.

5.1 Descriptive statistics

The first step consists of providing a rough overview of the data I use for my analyses. Table 3 lists the number of observations, the mean values, variances and value range of the emigration and the infrastructure variables. The values of the infrastructures refer to the arithmetic means for the five-year periods.

Table 3: Descriptive statistics of key variables

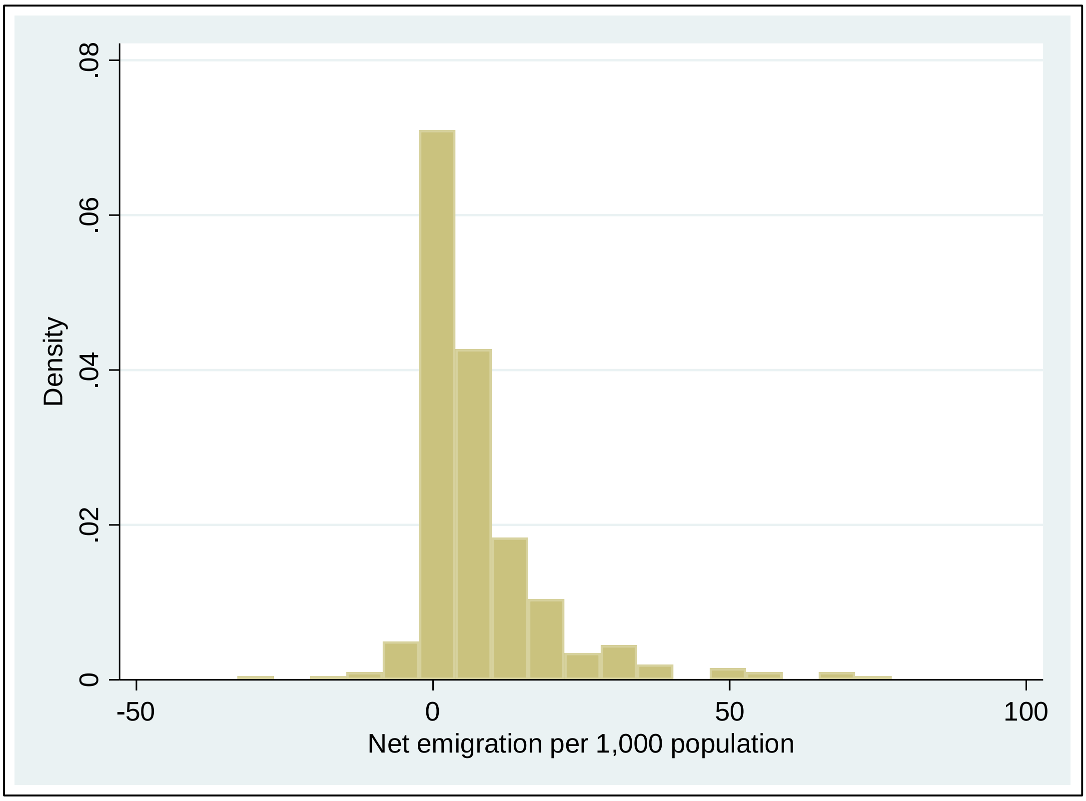
Variable	Observations	Mean value	Standard dev.	Min	Max
Emigration	329	7.9	12.0	-33.0	77.2
Road	329	3.9	1.2	1.5	6.7
Railroad	329	3.1	1.4	1.0	6.8
Electricity	329	4.5	1.5	1.3	6.9
Education	329	4.2	1.0	2.0	6.3
PLI	329	0.0	1.0	-1.8	2.4

Only observations with available data for all infrastructures and emigration are included, which explains the identical number of observations. As mentioned before, the 329 observations come from data for 120 countries. The third column reveals that the net emigration rate per country and year in the observation period amounts to roughly 8 per 1,000 population. The standard deviation is inflated to a value of 12 through a few observations with relatively extreme values, which is also

[27] To be precise, the exact variable is called “Gross Production Value” and is measured in constant international dollars.

reflected in the value range. The starkest outliers at the top of the distribution are Uruguay (77 in period 2), Venezuela (71 in period 3) and Bulgaria (68 in period 3), while Azerbaijan (-33 in period 1) is the most extreme case at the lower end of the distribution. Nevertheless, the histogram[28] in figure 5 reveals that the bulk of the observations is around or slightly above zero. The following regression analyses were also conducted under exclusion of these extreme observations, which did not change the results in any relevant way.

Figure 5: Distribution of net emigration rate



Source: Own representation based on data from (United Nations Department of Economic and Social Affairs, 2021).

The picture for the infrastructure variables looks quite different. Due to the data collection method, all values are limited to a range from 1 to 7, which makes stark outliers impossible. Mean values are between 3.2 (Railroad) and 4.5 (Electricity) while the standard deviations cover values between 1 and 1.5. The histograms in figures A-8 to A-12 show the distributions of the four infrastructures variables and the PLI index graphically. As a result of the standardization, the PLI index takes a mean value of 0, whereas the std. deviation amounts to 1. I proceed by examining how these variables relate to one another.

5.2 Bivariate relationships

The investigation of bivariate associations is often a useful starting point to approach the analysis of statistical relationships. Table 4 provides the correlation coefficients of the relationships

[28] Histograms are graphs that indicate the density of observations within a certain value interval (Agresti and Finlay, 2008, p. 34).

between emigration rates and infrastructures.[29] Once again, the five-year averages are used for the computations, which means that up to three observations per country are included.

Table 4: Bivariate correlations of emigration and infrastructures

	Emigration	Road	Railroad	Electricity	Education	PLI
Emigration	1					
Road	-0.08	1				
Railroad	-0.22*	0.75*	1			
Electricity	-0.05	0.74*	0.70*	1		
Education	-0.03	0.67*	0.72*	0.84*	1	
PLI	-0.13*	0.92*	0.90*	0.90*	0.83*	1

* $p<0.05$

The second column of the table reveals modest to weak negative relationships between the infrastructure variables and emigration. The correlation between emigration and railroad is the highest in absolute terms (-0.22) and significant. So is the negative relationship between emigration and the PLI index. Further, the table uncovers strong correlations among the infrastructure variables, which is little surprising since national income is a strong predictor for infrastructure quality of all kinds. This insight is important for the subsequent empirical analysis due to potential issues of imperfect multicollinearity.[30] Because of this data structure, I primarily use the outlined PLI index for the subsequent analyses, which also facilitates the interpretation. Nonetheless, the relationships between emigration and the individual coefficients will also be briefly discussed at the end of the chapter.

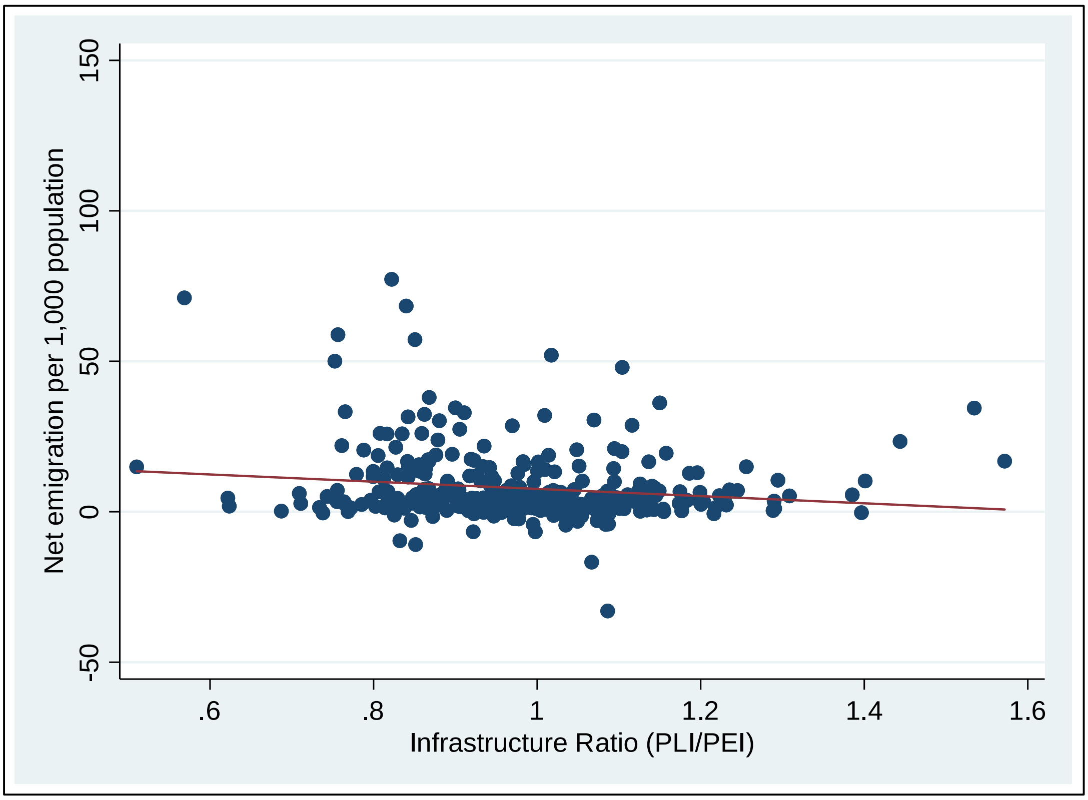
Considering the strong associations between the infrastructures, figure 6[31] plots emigration against the ratio of PLI and education, i. e., PEI, such that the differences between them can be better exploited. The intuition is to reveal whether countries with a relatively high ratio (relatively good PLI compared to education infrastructure) show lower emigration rates than countries with lower ratios.

[29] This coefficient measures the strength of the linear relationship between two variables. It ranges from -1 (perfect negative relationship) to 1 (perfect positive relationship) (Lane, 2003, p. 170).

[30] The issue of imperfect multicollinearity arises when regressors are highly correlated with another. Unlike perfect multicollinearity, it does not prevent the estimation of the regression, but it decreases the precision of the estimates (Stock and Watson, 2006).

[31] For illustration purposes, the PLI variable was adjusted such that both PLI and PEI have mean values of 4 (and standard deviations of 1 as before).

Figure 6: Emigration and infrastructure ratio



Source: Own compilation based on data from UN DESA (2021) and WEF (2021).

A slightly negative bivariate relationship can be observed which is in line with the expectations. The corresponding correlation coefficient is -0.15 and significant on a five percent level.

All the graphs and coefficients portrayed so far are merely bivariate correlations that help to provide a first impression but have little to do with the identification of causal relationships which is my actual research interest. To gain a deeper understanding of the relationships, a variety of further statistical analyses are conducted, all of which are based on the application of multivariate regression models of some form. The following section provides an overview of this type of regression models.

5.3 Multivariate regressions

Multivariate or multiple regression analyses offer two principal advantages compared to mere bivariate associations. First, they help increase the explanatory power of a model by incorporating multiple relevant determinants. Perhaps even more importantly, they allow to disentangle the various effects of different regressors which helps detect spurious correlations (Webster, 2012, p. 120). When an outcome variable is regressed on a variety of independent variables, the estimated slope coefficients can be interpreted as the change in the outcome variable given a one-unit increase of the respective independent variable when all other independent variables are kept constant (Stock and Watson, 2006, p. 193). Chapter 5.3.1 elaborates a bit further on the underlying methodology, before results are presented and discussed in the subsequent sections.

5.3.1 Methodology of multivariate regression analysis

A central underlying concept for the estimation of the coefficients is the ordinary least squares (OLS) method. The general idea is that the predicted values are as close to the observed values as possible. To this end, the parameter estimation minimizes the sum of the squared differences between predictions and observations (Stock and Watson, 2006, p. 118). Multivariate regressions with OLS estimators have become a popular tool for many statistical applications in economics and the social sciences because of some highly desirable properties. Oftentimes, these are summarized in the word "BLUE", which stands for "best linear unbiased estimator". Linearity means that the model is linear in its *parameters*, which does not necessarily mean that the dependent variable is a linear function of the independent *variables*.[32] A linear relationship is often viewed as a decent approximation, even if the true association is not expected to exactly follow a straight line (Agresti and Finlay, 2008, p. 272). The desired properties of OLS estimates are listed here (Webster, 2012, p. 46):

- ➔ Unbiased: The expected value of the estimation equals the true value of the parameter.
- ➔ Best: The estimates show a minimum variance around the true parametric value.
- ➔ Consistent: An increase in the sample decreases the variance, which brings the estimates closer to the true value of the parameter.
- ➔ Normally distributed: The estimates follow a normal distribution with a given variance around the parametric value.

For an estimator to have these properties though, a number of assumptions must be satisfied which are mostly known as the Gauss-Markov assumptions (Webster, 2012, p. 43-46; Stock and Watson, 2006)):

1) The model is linear in the parameters.

2) The sample of observations is from a random sample (i. e., the data is independent and identically distributed (i.i.d.)).

3) No independent variable is a linear function of any other independent variable(s) or constant (i. e., no perfect multicollinearity).

4) The error has an expected value of zero given any values of the independent variables.

5) The error has the same variance given any values of the explanatory variables (i. e., no heteroskedasticity).

Assumptions 1) – 4) must be met for the OLS estimates to be unbiased. If assumption 5) is additionally met, then the OLS estimates qualify as "BLUE". The linearity aspect has been explained above. If the empirical association follows a substantially different form, this can often be revealed graphically or when analyzing the residuals.[33] With respect to assumption 2), it must be noted that I do not work with a sample. Rather, I try to cover the population as well as possible given limitations due to data availability. I address a concern about a potential violation of the i.i.d. assumption in the following subchapter. Perfect multicollinearity (assumption 3)) constitutes a logical error which would be self-revealing, as models with such errors cannot be estimated. However, the close associations between many of the included variables (infrastructures, GNI,

[32] A regression model can be non-linear in the independent variables if, for instance, higher polynomial or logarithmic terms are included.

[33] The *residuals* are the deviations of the observations from the predictions of the model with *estimated* parameters, while the *error* expression is reserved for deviations of the observations from the model that uses the "true" but unknown parameters (Agresti and Finlay, 2008, p. 276).

freedom scores, etc.), imply problems with *imperfect* multicollinearity, which explains the use of the PLI index.

Assumption 4) is often the Achilles heel for identification strategies, and frequently caused by omitted variables. The correlation of an independent variable with an omitted variable that is at the same time a determinant of the outcome variable, leads to a correlation of this independent variable and the error term. This violation of assumption 4) produces a bias in the coefficient estimate (Stock and Watson, 2006, p. 186). To tackle the problem, one includes (if possible) such third variables in the regression model, such that the correlations are disentangled (Agresti and Finlay, 2008, p. 397 f.). The application of heteroskedasticity-robust standard errors is a relatively easy tool to (partly) relax assumption 5). Therefore, I run all following regression models using this standard error type. Generally, an analysis of the residuals is often helpful to detect signs for violations of assumptions 4) and 5).

These methodological remarks are necessary to understand and properly interpret the results of the following regression analyses.

5.3.2 Panel data regression analysis

Since my data is in panel structure, I start with regression methods that are appropriate for this context. Generally, there are three possibilities to run multivariate regressions with a panel data set. The first consists of simply pooling the data and then conducting an OLS regression. The main advantage compared to a cross-sectional design are the additional observations available, whereas the drawback of this approach is that the panel structure is not taken into account. Instead, all data points are viewed as independent observations, based on the (often unrealistic) assumption that no relevant time- or entity-specific effects exist. If this assumption is not met, the error terms are autocorrelated, representing a violation of the second Gauss-Markov condition. As a result, the estimates are biased and inconsistent (Xu, Lee and Eom, 2008, p. 578).[34] There are two alternative models that consider the panel structure of the data and are therefore usually preferred: the random effects (re) and the fixed effects (fe) model. The models differ in their assumptions on the error term, which can be divided into two components: an entity-specific error term (also called unobserved heterogeneity) and an idiosyncratic error term. In the re model, entity-specific effects are viewed as random variables, whereas the fe design treats them as coefficients to be estimated. In other words, the difference is whether parts of the error term are correlated with the regressors in the model. If that is the case, the re model produces biased estimates and the fe model should be preferred. If there is no correlation with the error term though, the re model would be the first choice given its higher efficiency due to the fact that it uses all variations in the independent variables (Xu, Lee and Eom, 2008, p. 582).

The appropriateness of the re or fe model can be checked by applying the Hausman test (see Xu, Lee and Eom, 2008, p. 584). The test's null hypothesis is that the error term is random, which is why the re model should be preferred if the null hypothesis cannot be rejected. For the given data though, it can be rejected, which is evidence in favor of a model with fixed effects. The intuitive explanation for the test result is that non-randomness in the error term may result from spatial

[34] An estimation is inconsistent if the estimate is biased and the bias is resistant even to increases of the sample size (Stock and Watson, 2006, p. 422).

correlation, i. e., the possibility that the error term of a country in one period is not independent of the country's error term in the other periods.

In fe models, the influences of all country-specific characteristics that are stable over time are controlled for (e. g., cultural or geographic aspects), which is very helpful to lower the risk of omitted variable bias. In addition to country fixed effects, I also introduce time fixed effects that control for factors that affect all countries similarly in a given time period but change over time (e. g., certain aspects of technological innovation). An intuitive understanding of fe models is that the estimates are computed using merely the within-country variations of each entity over time. For the estimates to be unbiased, it is assumed that once fixed effects and controls are included, the idiosyncratic error term is independent of the regressors (Stock and Watson, 2006, p. 356; Xu, Lee and Eom, 2008, p. 579). The results of the fe regressions are displayed in table 5 (fixed effects are not listed):[35]

Table 5: Regression table – fe model

	(1)	(2)	(3)	(4)	(5)
	Emigration	Emigration	Emigration	Emigration	Emigration
PLI	-3.595	-7.235	-5.615	-2.901	-28.85**
	(5.460)	(7.362)	(6.646)	(6.252)	(14.49)
PEI	7.084	8.901	12.15	7.762	8.764
	(6.127)	(8.253)	(9.061)	(8.449)	(8.322)
PLI*PEI					6.319**
					(3.169)
Conflict		-0.217	-0.328	-0.175	-0.152
		(0.321)	(0.407)	(0.197)	(0.199)
Freedom		2.851	2.445	-1.437	-0.819
		(5.570)	(5.568)	(3.201)	(3.236)
Homicides		-0.00235	-0.0238	0.136	0.0804
		(0.381)	(0.414)	(0.343)	(0.322)
Natural Disasters		-0.0177	-0.00600	-0.0148	-0.0192
		(0.0348)	(0.0332)	(0.0284)	(0.0295)
Median Age			0.618	0.554	0.0546
			(1.664)	(1.803)	(1.794)
Population Density			-0.181	-0.160	-0.117
			(0.117)	(0.111)	(0.114)
Urbanization			-0.347	-0.105	0.0253
			(0.718)	(0.698)	(0.721)
Migrant Stock			-0.411***	-0.378***	-0.342***
			(0.117)	(0.129)	(0.127)
GNI				-0.0118	-0.219
				(0.686)	(0.674)
Observations	329	258	252	237	237
R-squared	0.030	0.058	0.234	0.211	0.235
Number of Countries	120	105	103	99	99

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

[35] GNI is measured in 1,000 dollars.

The table shows five specifications which vary in the selection of control variables. The infrastructure variables are part of each specification and by adding controls I observe whether noteworthy changes regarding their coefficients occur. In specification (2), the variables to capture forms of forced migration, that I introduced above, are added. The variables median age, population density, urbanization and migrant stock are included in specification (3), before adopting GNI (4). The interaction term is only added in the fifth specification to facilitate the interpretation, but table A-3 shows the estimates of all other specifications when the interaction term is included.[36] This procedure of a gradual inclusion of controls serves to test the robustness of the findings and is kept for all following models.

It can be noted that the table shows only few significant results, which is further discussed below. The focus is first set on the signs and magnitudes. Regarding PLI, it can be observed that the estimate is consistently negative across all specifications which is in line with the prediction. The magnitudes suggest that an increase of PLI by one standard deviation is associated with a drop in emigration by approximately 3 to 7 (net) emigrants per 1,000 population. This seems quite substantial given that 50 percent of all emigration observations take values of between 1.8 and 10.2 per 1,000. For a large source country, such as Mexico with a population of roughly 125 million people, this could be translated into a decline of 375,000 to 875,000 emigrants. The fact that the inclusion of GNI reduces the magnitude may have to do with its particular role that was mentioned above. In specification (5), the magnitude (and sign) of PLI depends on the level of PEI quality. The coefficient for PLI is negative and indicated as significant, if PEI takes the value 0. However, given that the lowest PEI score is approximately 2, the coefficient is never statistically significant within the value range of PEI.

The sign of the PEI coefficient is consistently positive, which is in line with H2 and the respective literature (see brain drain section). The point estimates are even larger than the ones for PLI with a one standard deviation increase in PEI infrastructure being associated with an increase in emigration of between 7 and 12 out of 1,000 people. However, the estimated interaction term is also positive and even significant in specification (5), which was not expected. Table A-3 shows that the interaction coefficient is positive not only in specification (5) but also in all other specifications (even though it is then insignificant).

With respect to the control variables, it is most surprising that the coefficient of migrant stock has a highly significant negative sign, which stands in contradiction with the migration economics literature as discussed above. A possible econometric explanation for the finding is linked to the set-up of the empirical model. Since the migration flow variable is defined as the difference between the migrant stock in period *t+1* and the stock in period *t*, the latter is represented both on the left- as well as on the right-hand side of the regression equation. Assuming a systematic measurement error in the data, this structure would cause a downward bias in the coefficient. This bias may be particularly influential in the given fe model, where very little within-country estimation can be exploited for the identification. Further, also from a theoretical point of view this result may be more plausible than it first seems. Clemens (2014) explains that a negative

---

[36] The inclusion of an interaction term complicates the interpretation as it makes the coefficient of PLI dependent on the level of PEI and vice versa. Thus, the assessment of the coefficients' magnitudes, significances, and signs would become more complex.

correlation between emigration and economic development may be observed even for lower income countries if short-term effects are looked at, especially since the "emigration life cycle", i. e., the time it takes a country to move from the left to the right side of the inverted U-curve, occurs over many generations. He argues that if a country faces a temporary negative economic shock, parts of the population leave in the short term to look for better opportunities abroad. Once the country recovers though, many migrants return home. This differentiation between a long-term trend and short-term fluctuations may apply to migrant stocks as well. If only within-countries variations and relatively short time periods (5-year periods in this case) are considered, then temporary shocks may increase the outflows in the respective period, leading to an augmented migrant stock at the beginning of the following period. In the latter, the situation at origin improves again which incentivizes many of the temporary migrants to return home. The between-country differences on the other hand, reflect rather long-term trends, as they are the result of accumulated past migration flows. As I demonstrate below, in all other models, migrant stock has the expected significantly positive coefficient.

A similar case could be made for population density, which takes a negative (yet insignificant) sign here. While countries with higher population density might generally show higher levels of emigration (e. g. because of tight employment or housing markets), it is intuitive that in the short run the outflow of people lowers the density of the population at home. The variables controlling for forced migration appear to be of little relevance (lack of significance and low contribution of explanatory power). One relevant aspect in this context is that conflicts and natural disasters mainly lead to temporal and/or internal rather than international migration. Additionally, such extreme events may lead to a forceful displacement of some people, but they can also deprive others from the resources necessary for migration (Castelli, 2018; Food and Agriculture Organization of the United Nations et al., 2018). Even though the freedom variable was adopted as a control for forced migration, the literature suggests that freer societies are also associated with higher migration levels, which is in line with the negative coefficient in (4) and (5) (Ashby, 2010).[37]

To better understand the positive coefficient for median age, it is helpful to highlight that the average median age across all included countries is 28. While migrants are considered young in the eyes of most industrialized societies, the typical age in which migrants leave (late 20s or early 30s) is an above-average age in most principal sending countries. This suggests that countries with very low median ages still have the peak of their emigration rates ahead of them. Urbanization and GNI seem to play only minor roles in the model. Indeed, the adoption of GNI does not add explanatory power as the within R-squared values show. Overall, the model can explain up to 24 percent of the within-country variation.

Against the background of the mobility-transition curve literature and the policy debate about migration, it is of particular interest to look at the infrastructure coefficients for developing countries. Therefore, table A-4 uncovers the results of the fe regression under exclusion of all high-income countries as they are defined by the World Bank (2021c). The magnitudes of the PLI coefficients are now roughly twice the size they used to be, which is a remarkable change. Due to

[37] Remember that higher values indicate a lower degree of freedom.

the large standard errors, the coefficients still lack significance. On the other hand, the coefficient of PEI becomes very small, once GNI is controlled for.

Following the Gauss-Markov assumptions, a quick look must be taken at the distribution of the residuals. Figure A-13 plots the residuals of the specification (5) estimation against the fitted values.[38] The error terms do not look normally distributed which is also confirmed by a Jarque-Bera normality test.[39] The main issue with the distribution appears to be heteroskedasticity, since the variance of the residuals increases largely in the fitted values. This issue is addressed (yet possibly not fully resolved) by the application of heteroskedasticity-robust standard errors that are applied in all regressions. On the other hand, the graph does not raise much concern about systematic deviations from 0, which is desirable.[40]

Due to the non-normal distribution of the residuals but even more so because of the lack of significance for most coefficients, all the above findings must be interpreted with caution. In the absence of statistical significance, there is a substantial probability that deviations from 0 are merely coincidental. In this case however, there is a good explanation for why significant results could not be expected in the fe model, which is linked to the low number of periods included as well as to the low degree of variance per entity across these three periods. This aspect is highlighted in Table 6table 6 where the variation in the infrastructure variables is decomposed into its components. The table reveals that the bulk of the variation is between rather than within countries.

Table 6: Decomposition of variation in infrastructure variables

Variable		Standard Deviation
PLI	overall	0.90
	between	0.90
	within	0.16
PEI	overall	1.00
	between	1.00
	within	0.20

The standard deviations reveal that there is at least five times as much variation in the infrastructure variables between than within countries. Consequently, the data offers only little variation for within-country identification of the coefficients. This is a strong empirical argument to complement the fe model by the re model that also takes between-country differences into account despite the mentioned concerns about correlations between regressors and error term.

The results of the re regression model are presented in table 7 and follow the same structure as above. Table A-5 shows the results for all specifications under inclusion of the interaction term.

[38] Dummy variables are used here for incorporating country and time fixed effects.
[39] See Jarque (2011) for information on the test.
[40] The distributions of the residuals in the further models follow similar patterns but will not be commented on further.

Table 7: Regression table – re model

	(1)	(2)	(3)	(4)	(5)
	Emigration	Emigration	Emigration	Emigration	Emigration
PLI	-4.003**	-3.465*	-2.527	-0.657	-4.759
	(1.992)	(1.927)	(1.926)	(1.535)	(4.273)
PEI	2.807	1.939	-1.329	-3.720**	-3.068*
	(1.878)	(2.153)	(2.363)	(1.692)	(1.859)
PLI*PEI					0.954
					(0.892)
Conflict		-0.199	-0.130	0.0791	0.103
		(0.388)	(0.364)	(0.201)	(0.203)
Freedom		-0.970	-0.740	-1.137**	-1.082**
		(0.678)	(0.785)	(0.464)	(0.489)
Homicides		0.188**	0.204**	0.104*	0.111**
		(0.0944)	(0.0954)	(0.0550)	(0.0552)
Natural Disasters		0.0299	0.0350*	0.0406**	0.0418**
		(0.0219)	(0.0189)	(0.0184)	(0.0186)
Median Age			0.338	0.332*	0.351*
			(0.208)	(0.195)	(0.193)
Population Density			0.000478	0.000801	0.000609
			(0.000442)	(0.000560)	(0.000614)
Urbanization			-0.00613	0.0197	0.0286
			(0.0574)	(0.0608)	(0.0626)
Migrant Stock			0.0523***	0.0669***	0.0677***
			(0.0161)	(0.0145)	(0.0145)
GNI				-0.0478	-0.106
				(0.0920)	(0.108)
Constant	-3.536	1.521	0.499	10.70*	7.034
	(7.506)	(10.30)	(11.65)	(6.183)	(7.569)
Observations	329	258	252	237	237
Overall R-squared	0.033	0.082	0.21	0.271	0.274
Number of Countries	120	105	103	99	99

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Once again, all coefficients for PLI show a negative sign. While the coefficient is at least weakly significant (on the 5 or 10 percent level) in the first two specifications, significance vanishes when further controls are adopted. The point estimates of the specifications are clearly lower than in the fe model, especially when GNI is included.

Mixed results are found for the PEI variable. While a positive sign can be observed in the first two specifications, the sign changes in (3). In specification (4), the negative coefficient even becomes significant. With respect to specification (5) though, the indicated significance for a PLI value of 0, vanishes due to the interaction term for almost all of the observed PLI levels.[41] Similarly to the

[41] Significance on the 10 percent level prevails merely for the seven observations with the lowest PLI levels.

PLI coefficients, the magnitudes of the point estimates of the PEI coefficients are now much smaller than in the fe model.[42]

Table A-5 in the appendix provides extensive information about the interaction term. Its coefficient takes positive values at first, which become negative in (3) and (4). It is consistently insignificant and hardly adds explanatory power. The coefficient of migrant stock is now positive as expected and again highly significant. Further controls with mostly significant coefficients are homicide and median age, both of which show positive associations. A final aspect to be mentioned is that the adoption of GNI adds now substantial explanatory power to the model even though its coefficient does not become significant. In specifications (4) and (5) roughly 27 percent of the variation in migration flows can be explained through the model.[43]

As mentioned above, the third option for panel data analysis is the pooled regression model that treats all data points as independent observations. The Lagrange multiplier test investigates the appropriateness of a pooled OLS model (see Xu, Lee and Eom, 2008, p. 579). Its result justifies the application of a pooled OLS model given my data structure. Since the results are very similar to the ones of the re model, the regression tables are printed in the appendix (table A-6), but not discussed here.

At this point I would also like to briefly touch on the individual roles of the PLI index components. As argued above, they were summarized into one index due to issues of imperfect multicollinearity and to facilitate the interpretation.[44] However, relevant differences in their impacts are plausible, for instance because they may influence migration costs to different degrees. Regression tables with the PLI components for the various models (including cross-country) are shown in tables A-7 to A-10. The results imply that the strongest driver for the negative PLI coefficient are railroads. Road quality, on the other hand, rather appears to have a slightly positive (yet insignificant) association with emigration. The electricity coefficient is mostly negative and insignificant.

5.3.3 Cross-country regression analysis

Besides the presented panel data analyses, a transformation into a pure cross-sectional data structure is feasible by building the averages across the entire 15-year period. While this transformation may first seem counterintuitive because of the highlighted advantages of panel data, it can add additional insights in this context as it allows for the analysis of more long-term migration flows. Referring back to the argument made by Clemens (2014) regarding the impact of different time horizons on the association between national income and emigration, it is possible that similar patterns apply to the effects of infrastructures. Hence, short-term analyses might be more strongly influenced by temporary fluctuations. For instance, temporary crises could drive up emigration rates and at the same time constrain the means for infrastructure investments. Alternatively, in the case of counter-cyclical economic policies, infrastructure projects might be realized exactly in times of crises, in which case the improvement of infrastructure quality would happen to coincide with larger outflows of people. Such shocks may take different forms and if

[42] The magnitude ranges from an increase of 2.8 people per 1,000 inhabitants for a one-standard deviation increase in education infrastructure in (2) to a decrease of a slightly larger magnitude in (4).
[43] The exclusion of high-income countries does not yield significant changes this time and is therefore not further discussed.
[44] The variance inflation factor for some of the infrastructures (particularly education) can take values of up to 10 when all infrastructures are included separately.

they are not adequately captured by the control variables, patterns like this could affect my estimates in short-term analyses. Generally, the short-term management of migration flows tends to fall under the domain of regulatory policies (Ortega and Peri, 2013). The effects of infrastructures on the other hand are likely to unfold over longer time periods, and therefore a cross-country design is viewed as a sensible complement to investigate more sustained flows.

Since this part of the analysis uses the variables' average values for the entire observation period, the data follows no longer a panel structure. The regression equation changes to:

$$\begin{aligned} Emigration_i = \beta_o &+ \beta_1 PLI_i + \beta_2 PEI_i + \beta_3 (PLI_i * PEI_i) + \beta_4 Conflict_i + \beta_5 Freedom_i \\ &+ \beta_6 Homicides_i + \beta_7 NaturalDisasters_i + \beta_8 MedianAge_i \\ &+ \beta_9 PopulationDensity_i + \beta_{10} Urbanizaiton_i + \beta_{11} MigrantStock_i \\ &+ \beta_{12} GNI_i + u_i \end{aligned}$$

Table 8 shows the results for an OLS multivariate regression analysis.

Table 8: Regression table – cross-country model

	(1)	(2)	(3)	(4)	(5)
	Emigration	Emigration	Emigration	Emigration	Emigration
PLI	-4.622**	-3.668*	-2.304	-1.369	-0.733
	(1.892)	(1.938)	(1.768)	(1.913)	(4.452)
PEI	3.180*	2.222	-0.768	-1.039	-0.592
	(1.697)	(1.780)	(2.314)	(2.269)	(2.922)
PLI*PEI					-0.196
					(1.026)
Freedom		-0.689	-0.623	-0.972	-0.988
		(0.607)	(0.717)	(0.701)	(0.709)
Conflict		-0.0944	0.0547	0.0844	0.0761
		(0.510)	(0.388)	(0.342)	(0.349)
Homicides		0.208***	0.176**	0.119*	0.118*
		(0.0738)	(0.0725)	(0.0651)	(0.0659)
Natural Disasters		$-6.95*10^{-8}$	$5.11*10^{-9}$	$7.99*10^{-9}$	$7.99*10^{-9}$
		$5.08*10^{-8}$	$4.07*10^{-8}$	$4.95*10^{-8}$	$5.01*10^{-8}$
Median Age			0.198	0.150	0.147
			(0.208)	(0.207)	(0.210)
Population Density			0.0009**	0.0012**	0.0012**
			(0.0004)	(0.0005)	(0.0005)
Urbanization			-0.00282	-0.0108	-0.0124
			(0.0589)	(0.0614)	(0.0624)
Migrant Stock			0.0662***	0.0705***	0.0703***
			(0.0150)	(0.0143)	(0.0143)
GNI				-0.0364	-0.0235
				(0.0996)	(0.117)
Constant	12.19***	13.18***	7.181	8.676	7.377
	(2.427)	(4.916)	(5.161)	(6.210)	(9.682)
Observations	120	106	104	103	103
Adjusted R-squared	0.051	0.104	0.400	0.400	0.394

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

The coefficients for PLI are roughly in line with the above findings. The sign is once again consistently negative, and the magnitudes are similar to those in the re model/pooled OLS model, which is little surprising given that the estimates in these two models were largely determined by cross-country differences. The adoption of GNI lowers once more the (absolute) size of the coefficient. For specification (5), it must be noted that the interaction term is this time negative, which means that adding the interaction term increases the coefficient's magnitude for higher values of PEI infrastructure. However, also the combination of coefficients does not become significant (not even for the highest scores of PEI infrastructure).

The analysis of PEI resembles the picture from the re model too. Once more, the changing sign can be observed. The difference is that the negative coefficient in the cross-country model never becomes significant, which remains true if accounting for the interaction term in the last specification. All negative coefficients are quite small.

The most remarkable difference vis-à-vis the above results is arguably the negative sign of the interaction term. It is now consistently negative as table A-11 reveals. At the same time though, except for specification (1) it is not significant and in (3) and (4) very small. Thus, it would be hasty to derive from these results that the interaction term has a different impact in the long term compared to the short term.

While the control variables are little surprising, another noteworthy difference is that substantially more of the variance in the outcome variable can now be explained. The adjusted R-squared value amounts to 40 percent for specifications (3) to (5). Overall, this could indicate that the model is better at explaining longer term migration flow trends, while there is more "noise" in the short-term emigration patterns. This claim is also supported when comparing the regressions with the individual PLI components. The explanatory power of the infrastructure variables for the entire 15-year period in table A-10 is much higher than for the tables with panel data results A-7 to A-9.

5.3.4 Instrumental variable regression analysis

In this last section of the analysis part, I address the other major issue in terms of identification strategies, which is reverse causality. Similar to biases due to omitted variables, a correlation between regressors and the error term can also stem from simultaneous/reverse causality. This represents a violation of the Gauss-Markov assumptions and can result in biased and inconsistent estimations. To better illustrate the issue, it is sensible to differentiate between exogenous and endogenous regressors, i. e. regressors which are uncorrelated (exogenous) and correlated (endogenous) with the error term, respectively. A popular tool to eliminate the bias in the endogenous variables consists of finding one or multiple instruments that help isolate the unproblematic (uncorrelated) part of the variable. To this end, the instrument must meet the following two criteria: First, it must be sufficiently correlated with the endogenous variable (relevance) and, second, it should not affect the outcome variable through any channels other than through the endogenous variable (exclusion restriction). If other channels exist, they must be controlled for (Stock and Watson, 2006, p. 422).

Instrumental variable regressions often take place through a two-stage-least-squares (TSLS) approach. In the first stage, the endogenous variable[45] is regressed on the instrument(s) as well as

[45] For the sake of simplicity, I assume here that just one endogenous variable exists.

on all exogenous variables of the regression model. This first stage regression is used to predict uncorrelated values for the endogenous variables. These values are applied in the second stage, in which the outcome variable is regressed on all exogenous variables as well as the predicted values of the endogenous variable, yielding (if the assumptions are met) unbiased estimates. Instrumental variable regressions require at least as many instruments as endogenous regressors. A model is called overidentified if there are more instruments than endogenous variables, exactly identified if the numbers are identical or underidentified if there is an excess of endogenous regressors, in which case the TSLS method cannot be conducted (Stock and Watson, 2006, p. 434, f.)

With regard to the infrastructure-migration relationship, a potential source of reverse causality are remittances. Countries with high shares of emigrants are likely to benefit from substantial inflows of remittances that can be used to maintain or expand infrastructures (IOM, 2019a). On the other hand, countries with high emigration rates might see increasing shares of their territories becoming abandoned which could discourage policymakers from investments in local infrastructures. The above remarks on the brain gain mechanisms demonstrate that emigration rates influence the demand for PEI. These examples show that reverse causality may pose a threat to a proper identification, but the sign of the bias is not evident ex ante.

The search for a suitable instrument is a difficult endeavor. Geographic features are quite popular as they typically rule out the possibility of reverse causality (Grossmann and Osikominu, 2019). At the same time, geographic factors can matter for the provision of PLI as the costs for building roads, railroads or electricity lines evidently depend on geography (Rodrigue, 2020). Different geographic features, such as the ruggedness index, have been tested. The candidate that has turned out to be sufficiently relevant is the share of wetland in a country. The rationale is that wetland complicates the installation of infrastructures, such as transport lines or electricity grids, which is why – all other things equal – it can be expected to be negatively correlated with PLI. Unfortunately, an instrument for PEI could not be found. The wetland data is obtained from the OECD (2021) database on land cover and provides information on almost all countries included in this research.

The instrument comes with two important drawbacks. First, the wetland variable is constant over time, which is why it cannot be combined with the application of fixed effects. Thus, it is not possible to adjust for unobserved heterogeneity and reverse causality at the same time. Additionally, there is reason to question the exclusion restriction. While it is implausible to assume that emigration affects the share of wetland at origin in any relevant way (no reverse causality), the instrument could be associated with emigration through channels other than PLI, especially because emigration can be influenced by a large variety of factors. One threat is the link of wetland with different climate conditions which are in turn associated with agricultural production. A report of multiple international organizations from 2008 stresses the importance of food insecurity as a push factor of international migration, while at the same time emphasizing the role of climate conditions for food production (Food and Agriculture Organization of the United Nations et al., 2018). This may be an important channel, which is why I control for agricultural production in both stages. Links between hazardous climate events and emigration are already covered in the natural disasters control variable. These precautions can lower but not eliminate the possibility that the exclusion restriction is violated. As I am unaware of a better alternative though, wetland is selected for the TSLS approach. The first stage regression is based on this regression equation:

$$\widetilde{PLI_i} = \pi_0 + \pi_1 Conflict_i + \pi_2 Freedom_i + \pi_3 Homicides_i + \pi_4 NaturalDisasters_i + \pi_5 MedianAge_i + \pi_6 PopulationDensity_i + \pi_7 Urbanization_i + \pi_8 MigrantStock_i + \pi_9 GNI_i + \pi_{10} Agriculture_i + v_i$$

The predicted PLI values ($\widetilde{PLI_i}$) are then incorporated into the second regression equation:

$$Emigration_i = \beta_o + \beta_1 \widetilde{PLI_i} + \beta_2 Conflict_i + \beta_3 Freedom_i + \beta_4 Homicides_i + \beta_5 NaturalDisasters_i + \beta_6 MedianAge_i + \beta_7 PopulationDensity_i + \beta_8 Urbanization_i + \beta_8 MigrantStock_i + \beta_9 GNI_i + \beta_{10} Agriculture_i + u_i$$

The results of the first-stage as well as the second-stage regressions are shown in table 9.

Table 9: Regression table – instrumental variable model (TSLS)

	(1)	(2)
	PLI	Emigration
Wetland	-0.0688***	
	(0.0196)	
Freedom	0.0116	-0.788
	(0.0375)	(0.631)
Homicides	0.00407	0.138**
	(0.00375)	(0.0600)
Natural Disasters	$3.13*10^{-8}$*	$4.31*10^{-7}$**
	($1.59*10^{-8}$)	($2.17*10^{-7}$)
Conflict	0.00805	0.00377
	(0.0333)	(0.330)
Median Age	0.0201*	0.198
	(0.0109)	(0.209)
Population Density	-0.00013***	0.00097
	(0.00004)	(0.0006)
Urbanization	-0.00207	-0.00636
	(0.00402)	(0.0556)
Migrant Stock	-0.00099*	0.0640***
	(0.0006)	(0.0140)
GNI	0.0431***	-0.0439
	(0.00565)	(0.191)
Agricultural Production	$-1.78*10^{-9}$	$-4.05*10^{-8}$**
	($1.47*10^{-9}$)	($1.76*10^{-8}$)
PLI		-2.312
		(4.454)
Constant	1.876***	7.106
	(0.344)	(8.611)
Observations	103	103
Adjusted R-squared	0.748	0.408

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

In the first stage, the coefficient of wetland is found to be negative as expected and highly significant. The F-test in the first-stage regression, that tests the hypothesis whether the coefficient of the instrument is equal to zero yields a value of 13, which exceeds the necessary minimum of

10. However, a value of 13 value does not represent a very strong instrument which leads to more uncertainty in the second-stage estimation and a lower probability of significant results (Hanck et al., 2020).[46]

In the second stage, the focus is on the coefficient of $\widetilde{PLI_{\iota}}$. The sign remains negative, but the coefficient does not become significant, which may also have to do with the mentioned increase in uncertainty. The magnitude of the coefficient is slightly higher than it was in the other estimates under the inclusion of GNI as a control. Even though caution is required in the face of the large standard errors and uncertainties regarding the exclusion restriction, this result suggests that the above findings have not been substantially driven by reverse causality. Most other coefficients are as expected or of little relevance. The share of explained variation (around 40 percent) is very similar to the share of the cross-country model without instruments.[47]

When high-income countries are omitted, the instrumental variable regression even yields a significantly negative result for PLI as table A-12 highlights.[48] The point estimate of approximately -12 is also very large. This supports the hypothesis of a substantial negative association between PLI and emigration particularly for developing countries.

5.4 Summary of results

This chapter has collected extensive evidence on the empirical relationship between the examined infrastructures and emigration. As all designs have their strengths and weaknesses, the intention has been to cover a wide range of different approaches that contribute to an overall picture. This picture – imperfect as it is – implies the following conclusions:

First, it lends support to H1 (negative effect of PLI on emigration). The coefficients of the PLI index have been consistently negative and sometimes (though not often) significant. Possible explanations were suggested, when significance was lacking (e. g., little within-country variation or imprecision due to instrument). The negative sign was resilient to the application of fixed effects or an instrumental variable approach. While the magnitude of the coefficient appears substantial, the separated analysis of the infrastructure index implies that the main driver for this negative relationship is railroad infrastructure.

Second, the evidence regarding H2 (positive effect of PEI on emigration) is ambiguous. The coefficient has taken positive and very substantial values in the fe regression, but the picture changed in the following models. When many controls were added, the coefficient sign turned negative, and the magnitude declined largely. Unlike in the case of PLI, no suitable instrument could be found to control for reverse causality. One interpretation, that would be in line with most findings in the literature, would assume biases due to unobserved heterogeneity when fixed effects are removed. This would constitute a form of omitted variable bias, in the sense that PEI is correlated with one (or multiple) time-constant variables which have a negative effect on emigration rates. For instance, if PEI is of particularly high quality in countries with a culture that discourages people from leaving, this would generate a downward bias in the estimate for PEI. The fixed effects, capturing the cultural influence, would then reveal an unbiased estimate. In the

[46] The instrument did not meet the relevance criterion in the panel data models. Therefore, the TSLS application is limited to the cross-country analysis.
[47] It was expected that a drop in the agricultural output may cause people to leave.
[48] The relevance criterion is just met (F-test value of exactly 10).

face of the presented results however, H2 can clearly not be confirmed. The lack of support for H2 could also be related to the rejection of the sub-hypothesis about the interaction term. This one was expected to be negative, but the empirical analysis mainly revealed insignificantly positive coefficients.

Given the long-term nature of infrastructure quality, analyses for longer time horizons seem to yield more explanatory power. This is just one aspect that highlights that the findings must also be interpreted in the context of the selected research design(s), data availability, the processes of data collection etc. Therefore, a critical discussion of the research design and the results is presented in the following chapter.

6 Discussion of contributions and limitations

This section is structured into three parts: First, the specifications of the model design are critically reflected before turning to questions of data selection and data quality. Lastly, I address the limitations of the methods used in my empirical analysis.

6.1 Discussion of the model design

One key contribution of this work is the development of a novel classification of infrastructures, which allows a different perspective on the infrastructure-emigration relationship. I could also show that based on this framework, it is possible to formally derive clear hypotheses under the highlighted set of assumptions.

These assumptions serve to abstract from the highly complex reality, which is the key idea of a model. Naturally, the reduction of complexity comes at a cost because some aspects of people's behavior are neglected. For instance, I assumed that individuals take merely one migration decision in their lives and determined when this decision is taken (period *t*). Evidently the reality looks different.[49] However, these simplifications are unlikely to largely distort the mechanisms I intend to study as labor migration is the dominant migration form and infrastructure quality can plausibly matter for return decisions as well.

Another simplification is the exclusive focus of the individual on their own utility. Evidently, most parents care also about their children's life opportunities. Therefore, it is possible that good PEI does not only function as a push factor (search for highest returns), but also as a pull factor for parents. This may be an alternative (non-statistical) explanation for the mixed regression results regarding PEI.

Besides, the question can be raised to which degree education is representative of PEI. Due to the focus on education, I model PEI with an "end date" (end of *t-1*) for the delivery of the infrastructures' services, i. e., education. This is not a necessary feature of PEI (e. g., health infrastructure) and further research should therefore take a broader perspective if adequate data can be found.

Finally, it should be noted that I model the *intention* rather than the act of migration. The feasibility of migration aspirations has not been part of this research. Once more, if good data is accessible,

[49] For instance, studying abroad or return migration is both common today (Koser, 2016).

the even more direct link between infrastructures and the *willingness* to leave could be explored. This point builds the bridge to questions regarding data availability and quality.

6.2 Discussion of data availability and quality

As pointed out above, I compiled a wide range of high-quality data from a variety of sources to enable a rigorous analysis of the research question under inclusion of a relatively large set of controls. Migration and infrastructure data were obtained for 120 countries, with many developing economies among them. In most cases, I have drawn on data from international organizations and the GCI, although, if necessary, I have selected other reliable sources as complements. My method of deriving migration flows from the stock data leads generally to an underestimation of emigration flows, as the migrant stocks decrease not only as a consequence of return migration but also when people die over time. Nevertheless, this method should not lead so a systematic bias and is recommended and applied by leading researchers in the field (Clemens, 2020, 2014; Dao et al., 2018).

Unfortunately, data that reveals the emigration type (i. e., labor migration, humanitarian migration, etc.) or the importance of the two presented mechanisms (income and amenity value) is not available.[50] Further, I cannot identify the role of additional channels, such as migration costs. Individual-level data and interviews of migrants could cast more light on that matter.

The infrastructure data convince through the high degree of comparability (identical data collection methods), the comprehensiveness (questions relate to the infrastructure systems as a whole) and the global reach (including the developing world), Moreover, the GCI is an internationally widely recognized data source (Dijkstra et al., 2011). The main drawback is the limitation to a time period that is relatively short for the analysis of sustained migration flows. Consequently, I forewent the use of time-lagged variables and approximated infrastructure mean values in periods 1 and 3. Most importantly, the low amount of within-country variation impacted the efficiency of the fe model. Generally, concerns about biases due to surveys as a data collection method can be raised.[51]

6.3 Discussion of the statistical methods

Finally, the applied statistical methods need to be discussed. In this paper, I followed the objective of combining multiple approaches to collect a solid evidence base. This is especially desirable as all of the presented designs are associated with certain advantages and drawbacks, such that the most complete and differentiated picture could be painted by presenting and explaining multiple approaches.

The results lack significance for many estimated coefficients, which may have to do with the data structure or the statistical design, as stated above. Nevertheless, it is important to remember that insignificant coefficients have a non-negligible probability of being a product of chance, which was one reason to check for robustness across various models and specifications (Stock and Watson, 2006).

[50] While data for the number of refugees inside a country (i. e., immigrants) is provided by UN DESA (2021), this is not the case for emigrants.
[51] A summary of potential issues is printed in (Schumann, 2012).

A general problem of non-experimental analyses is that biases due to omitted variables cannot be ruled out, even though the inclusion of control variables and fixed effects serve to reduce this risk. The same is true for the issue of reverse causality, which I addressed by instrumenting PLI.

Another limitation concerns the functional form of the examined relationship. In the analysis, I assume an (approximately) linear relationship between infrastructures and emigration. As the “true” relationship may take all kinds of shapes, the analyses were also run under the inclusion of squared and logarithmic variables. In the absence of considerably better or different results with respect to measures of fit, significance, sign and magnitudes of the coefficients, the linear model was kept because it facilitates the interpretation. Besides, the linear form follows most directly from the theoretical model.

Finally, I need to stress that the relationships discovered on the global scale do not have to hold in each country individually, since particular national contexts can influence migration patterns.

To conclude, the examination of the research question is based on a novel theoretical framework, high-quality data sources, uses a variety of well-established research methods, and follows (if possible) the approach of leading researchers in the field. However, despite the mentioned efforts to reduce risks regarding my identification strategy, a number of considerable limitations remain and suggest a cautious interpretation of the findings. The results should also be seen as encouragement for further research that could address some of the highlighted concerns.

7 Conclusions and policy recommendations

The objective of this paper was the analysis of the relationship between infrastructures and emigration rates. At the beginning of the paper, I illuminated in broad strokes the current state of research on migration. In this context, I defined the key concepts, before giving an overview of relevant migration trends, which uncovered an increase in the volumes with a high degree of geographic heterogeneity. Migrant populations were found to be particularly large in higher-income countries. Regarding the composition of the migrant population, some patterns with respect to gender, age, and education of migrants were emphasized.

Subsequently, the opportunities and risks of migration were considered from different angles. The relocation of labor – at least at moderate levels – seems to have the potential to boost global productivity. Migration fosters innovation, trade relations and the diversity of goods. However, the challenges (e. g., negative impact on cooperation and the issue of brain drain) of migration were also pointed out.

The relationship between migration and development was investigated and the misconception of the “addressing the root causes of migration”-narrative, which builds on the standard neoclassical model, was discussed. Empirically, a correlation between development and emigration is observed that roughly follows the shape of an inverted U. This discrepancy between standard neoclassical predictions and observed migratory patterns represents a puzzle for which different explanations (e. g., credit constraints, urbanization, and demographic change) have been put forward. Yet, I emphasized the uncertainty regarding the effects of individual factors on emigration, particularly if these factors are closely linked with the overall process of economic development.

My work adds to this literature by examining the role of infrastructures, because only education has received significant attention in the international migration literature so far. I developed a classification of infrastructures that distinguishes between a place-based (PLI) and a people-based (PEI) type, which are expected to differ in how they (dis)incentivize migration. I used road, railroad, and electricity infrastructures to represent PLI, and education infrastructure as a proxy for PEI. On this basis, I derived the hypotheses that PLI affects emigration negatively, whereas PEI should exert a positive impact on emigration.

After the presentation of my data sources (e. g., UN DESA and the WEF's GCI) and a brief overview of descriptive statistics and bivariate correlations, I made use of various regression analysis techniques, such as fixed effects and an instrumental variable approach to test the hypotheses. I justified the applications of the different models, included a set of control variables, and interpreted the results. Even though significant results were rather rare, the consistently negative coefficient for the PLI index lent credence to H1, which is particularly true for developing countries. However, a more differentiated perspective implies that considerable differences between the PLI components exist. The findings for PEI were more mixed than expected, potential reasons for which have been discussed (e. g., unobserved heterogeneity and "pull effect"). Including an interaction term for the two variables did not prove to be fruitful. In chapter 6, I critically reflected on my findings and limitations linked to the set-up of the model, the data and the statistical tools that were applied. Against this background further research would be helpful to consolidate the results and gain more insights about the relevant mechanisms.

Research efforts with respect to PLI, could also focus on the increasingly important role of ICT infrastructures, which fall into the place-based category but, at the same time, largely facilitate exchanges with home communities. Therefore, ICT may also encourage emigration through a different channel.

Furthermore, it would be desirable to gain insights into the extent to which certain population groups are affected differently by various infrastructures. For this and other reasons (more knowledge on the channels and regional contexts, etc.), individual-level data and the conduction of interviews could serve to deepen the understanding of the respective relationships.

The findings of my research can help understand and potentially steer labor migration in desired directions. This applies to both origin and destination countries and – given the substantial benefits of international migration – should not be limited to a one-sided effort of reducing migration numbers. Therefore, one recommendation consists of carefully evaluating (in an objective and evidence-based manner) opportunities and risks of different volumes of (various kinds of) emigration and immigration. Such an assessment could be translated into target value ranges that may then be incorporated into a comprehensive strategy.

Infrastructure development – be it in the own country or in the context of development cooperation – could then be one (out of multiple) instruments to work towards these targets. A key message of this research is that governments that face large-scale outflows of younger people do not need to respond by cutting education funding (see brain drain section) but can instead augment investments that generate benefits for staying. Required funds may be provided by destination countries with a desire to ease immigration pressures. In light of the heterogenous impacts of PLI components, a differentiated approach is recommended. Moreover, the strategic placement of PLI within a country should also be considered, because, for example, the process of urbanization may be influenced (Kalu et al., 2014).

As stated above, the results with respect to PEI infrastructure were not as clear-cut and would require further investigation. Essentially, two possibilities can be discussed. First, PEI may not be a significant driver of net emigration, possibly because good schools (and perhaps also hospitals, etc.) are not only a push, but also a pull factor, by attracting parents concerned about their children's education. The alternative interpretation, which is more in line with the current state of the literature, is that the absence of positive coefficients for PEI (in specifications with many controls) is due to unobserved heterogeneity, as the fe model implies. In either case, cutting education funding would certainly create highly undesirable effects, as education does not only drive productivity but also generates highly positive spillovers as discussed above. In countries with massive brain drain issues, specific regulations that create incentives to stay for those profession groups whose skills are urgently needed, may be a better alternative.

Receiving countries should be aware that education infrastructures in origin countries do not only shape the volumes but also the characteristics of immigrants that arrive. This influence on the kind of migration is important to consider as high-skilled migrants are often more desirable from an economic point of view but also more accepted in the host societies. However, such investments should not be sold to the public as a policy to reduce immigrant numbers. At the same time, developed countries should be aware of their responsibilities towards the lower income countries and refrain from aiming for the skills that are particularly scarce in the developing world.

This paper contributes to a better understanding of one of the most important issues societies around the world are facing and whose relevance is expected to grow further. I propose a new way to think about the infrastructure-emigration relationship and provide extensive empirical evidence that encourages further research. In a broader context, this work highlights the complexity and multifaceted nature of migration, which should make people wary of frequently propagated "simple solutions". In that sense, I contribute to the rigorous analysis that Collier (2013, p. 12) demands as a foundation on which to build a fruitful policy discussion.

Bibliography

Abbasa, Q. and Foreman-Peck, J. (2007). *Human Capital and Economic Growth: Pakistan, 1960-2003*. The Lahore Journal of Economics, 13 (1), pp. 1–27.

Adhikari, S. and Gentilini, U. (2018). *Should I Stay or Should I Go: Do Cash Transfers Affect Migration?* The World Bank Policy Research Working Papers 8525.

Agresti, A. and Finlay, B. (2008). *Statistical Methods for the Social Sciences*. 4th ed. Upper Saddle River, N.J: Prentice Hall.

Albouy, D., Cho, H. and Shappo, M. (2019). *Immigration and the Pursuit of Amenities*. Journal of Regional Science, 61(1), pp. 5–29.

Alesina, A. and La Ferrara, E. (2000). *Participation in Heterogeneous Communities*. The Quarterly Journal of Economics, 115 (3), pp. 847–904. doi: 10.1162/003355300554935.

Algan, Y., Hémet, C. and Laitin, D. (2011). *Diversity and Public Goods: A Natural Experiment with Exogenous Residential Allocation.* Institute for the Study of Labor (IZA), IZA Discussion Papers (No 6053).

Ashby, N. J. (2010). *Freedom and International Migration*, Southern Economic Journal, 77 (1), pp. 49–62.

Batista, C., Lacuesta, A. and Vicente, P. C. (2012). *Testing the "brain gain" hypothesis: Micro evidence from Cape Verde.* Journal of Development Economics, 97 (1), pp. 32–45.

Batista, C. and Vicente, P. C. (2011). *Do Migrants Improve Governance at Home? Evidence from a Voting Experiment*, The World Bank Economic Review, 25 (1), pp. 77–104.

Benhabib, J. and Jovanovic, B. (2012). *Optimal Migration: A World Perspective.* International Economic Review. 53(2), pp. 321–348.

Biffl, G. (2019). *Economics and Migration*. In: Inglis, C., Li, W., and Khadria, B. (Eds.). *Sage Handbook of International Migration*. Los Angeles: SAGE Publications.

Borjas, G. J. (2015). *Immigration and Globalization: A Review Essay*. Journal of Economic Literature, 53 (4), pp. 961–974.

Brock, G. and Blake, M. (2015). *Debating Brain Drain: May Governments Restrict Emigration?* Cary, United States: Oxford University Press.

Castelli, F. (2018). *Drivers of migration: why do people move?.* Journal of Travel Medicine, 25 (1).

Centre for Research on the Epidemiology of Disasters (2021). *EM-DAT Public*. Available at: https://public.emdat.be/data [Accessed date: 23/04/2021].

Chauvet, L. and Mercier, M. (2014). *Do return migrants transfer political norms to their origin country? Evidence from Mali.* Journal of Comparative Economics, 42 (3), pp. 630–651.

Clemens, M. (2014). *Does Development Reduce Migration?* Institute of Labor Economics (IZA). IZA Discussion Papers (No 8592).

Clemens, M. (2020). *The Emigration Life Cycle: How Development Shapes Emigration from Poor Countries*. IZA Discussion Paper 13614.

Clemens, M. and Postel, H. (2018). *Deterring Emigration with Foreign Aid: An Overview of Evidence from Low-Income Countries.* Population and Development Review, 44 (4), pp. 667–693.

Clemens, M. and Pritchett, L. (2019). *The new economic case for migration restrictions: An assessment*. Journal of Development Economics, 138, pp. 153–164.

Clemens, M. and Sandfur, J. (2014). *Let the People Go: The problem with strict migration limits.* Foreign Affair, (93 (1), pp. 152–159.

Clist, P. and Restelli, G. (2020). *Development aid and international migration to Italy: Does aid reduce irregular flows?* The World Economy.

Collier, P. (2013). *Exodus: How Migration Is Changing Our World.* Reprint Edition. Oxford: Oxford University Press.

Crawley, H. (2016). *Managing the Unmanageable? Understanding Europe's Response to the Migration "Crisis"*. Human Geography, 9 (2), pp. 13–23.

Czaika, M. and de Haas, H. (2012). *The Role of Internal and International Relative Deprivation in Global Migration.* Oxford Development Studies, 40 (4), pp. 423–442.

Dao, T. H. et al. (2018). *Migration and Development: Dissecting the Anatomy of the Mobility Transition.* Journal of Development Economics, 132 (C), pp. 88–101.

Dijkstra, L., Annoni, P. and Kozovska, K. (2011). *A new regional competitiveness index: Theory, methods and finings*. European Commission, Directorate-General for Regional Policy. Available at: https://ec.europa.eu/regional_policy/sources/docgener/work/2011_02_competitiveness.pdf [Accessed 29/04/2021].

Docquier, F. and Marfouk, A. (2006). *International Migration by Education Attainment, 1990-2000*. In: Özden, Ç. and Schiff, M. (Eds.). *International Migration, Remittances, and the Brain Drain*. Washington D.C., United Nations: World Bank and Palgrave Macmillian, pp. 151–200.

Docquier, F. and Rapoport, H. (2012). *Globalization, Brain Drain, and Development*. Journal of Economic Literature, 50 (3), pp. 681–730.

Dustmann, C., Fadlon, I. and Weiss, Y. (2011). *Return migration, human capital accumulation and the brain drain.* Journal of development economics, 95 (1), pp. 58–67.

Feliciano, C. (2005). *Educational selectivity in U.S. Immigration: How do immigrants compare to those left behind?* Demography, 42 (1), pp. 131–152.

Fisman, R. and Miguel, E. (2007). *Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets.* Journal of Political Economy, 115 (6), pp. 1020–1048.

Food and Agriculture Organization of the United Nations *et al.* (2018). *The linkages between migration, agriculture, food security and rural development.* Technical report. Rome. Available at: http://www.fao.org/3/ca0922en/ca0922en.pdf [Accessed: 19 April 2021].

Fourie, J. (2006). *Economic Infrastructure: A Review of Definitions, Theory and Empirics*. South African Journal of Economics, 74 (3), pp. 530–556.

Freedom House (2021). *Freedom in the World*. Available at: https://freedomhouse.org/report/freedom-world [Accessed: 3 April 2021].

Fried, S. and Lagakos, D. (2017). *Rural electrification, migration and structural transformation: Evidence from Ethiopia.* Regional Science and Urban Economics (103625).

Gould, J. D. (1979). *European inter-continental emigration 1815-1914: pattern and causes.* The Journal of European Economic History, 8 (3), pp. 593-679.

Grogger, J. and Hanson, G. H. (2011). *Income maximization and the selection and sorting of international migrants.* Journal of Development Economics, 95 (1), pp. 42–57.

Grossmann, V. and Osikominu, A. (2019). *Let the Data Speak? On the Importance of Theory-Based Instrumental Variable Estimations*. German Economic Review, 20 (4), pp. 831–851.

de Haas, H. (2007a). *International migration, remittances and development: myths and facts*. Third World Quarterly, 26 (8), pp. 1269–1284.

de Haas, H. (2007b). *Turning the Tide? Why Development Will Not Stop Migration*. Development and change, 38 (5), pp. 819–841.

de Haas, H., Castles, S. and Miller, M. J. (2020). *The Age of Migration: International Population Movements in the Modern World*. 6th ed. London: Red Globe Press.

Hamilton, B. and Whalley, J. (1984). *Efficiency and distributional implications of global restrictions on labour mobility: Calculations and policy implications*. Journal of Development Economics, 14 (1), pp. 61–75.

Hanck, C., Arnold, M., Gerber, A. and Schmelzer, M. (2020). *Introduction to Econometrics with R*. Essen: University of Duisburg-Essen.

Hatton, T. J. and Williamson, J. G. (1994). *What Drove the Mass Migrations from Europe in the Late Nineteenth Century?* Population and Development Review, 20 (3), pp. 533–559.

Hatton, T. J. and Williamson, J. G. (2005). *What Fundamentals Drive World Migration*. In: Borjas, G. J. and Crisp, J. (Eds.). *Poverty, International Migration and Asylum*. London: Palgrave Macmillan UK, pp. 15–38.

Helliwell, J. F., Layard, R. and Sachs, J. (2018). *World Happiness Report 2018*. New York: Sustainable Development Solutions Network.

Inglis, C., Li, W. and Khadria, B. (2019). *The SAGE Handbook of International Migration.* SAGE Publications.

Institute for Health Metrics and Evaluation (2021). *GBD Results Tool | GHDx*. Available at: http://ghdx.healthdata.org/gbd-results-tool [Accessed: 23/04/2021].

International Organization for Migration (2019a). *World Migration Report 2020*. Geneva: International Organization for Migration.

International Organization for Migration (2019b). *Glossary on Migration*. Available at: https://publications.iom.int/system/files/pdf/iml_34_glossary.pdf [Accessed: 3 October 2021].

Jarque, C. (2011). *Jarque-Bera Test*. In: Lovric, M. (Ed.). *International Encyclopedia of Statistical Science*. Berlin, Heidelberg: Springer, pp. 701–702.

Kalu, O., Ibiam J., Stephen, N. and Ijeoma, U. (2014). *Rural Infrastructure Development: A Tool for Resolving Urbanization Crisis. A Case Study of South-Eastern Nigeria*. World Environment, 4 (1), pp. 14-21.

Karadja, M. and Prawitz, E. (2018). *Exit, Voice, and Political Change: Evidence from Swedish Mass Migration to the United States*. Journal of Political Economy, 127 (4), pp. 1864–1925.

Kennan, J. (2012). *Open Borders*. National Bureau of Economic Research, Working Paper No 18307.

Klein, P. and Ventura, G. (2009). *Productivity Differences and the Dynamic Effects of Labor Movements.* Journal of Monetary Economics, 56 (8), pp. 1059–1073.

Koser, K. (2016). *International Migration: A Very Short Introduction, International Migration: A Very Short Introduction*. Oxford: Oxford University Press.

Lall, S. V., Timmins, C. and Yu, S. (2009). *Connecting Lagging And Leading Regions: The Role Of Labor Mobility*. The World Bank. Policy Research Working Paper No 4843.

Lane, D. (2003). *Introduction to Statistics, Open Textbook Library*. Available at: https://open.umn.edu/opentextbooks/textbooks/459 [Accessed: 10 April 2021].

Lucas, R. (2019). *Migration and development*: *The Role for Development Aid*. Migration Studies Delegation, 2019 (5).

Luttmer, E. F. P. (2001). *Group Loyalty and the Taste for Redistribution.* Journal of Political Economy. 109 (3), pp. 500–528.

Martin, P. L. (1993). *Trade and Migration: The Case of NAFTA.* Asian and Pacific Migration Journal, 2 (3), pp. 329–367.

Massey, D., Arango, J., Hugo, G., Kouaouci, A., Pelegrino, A., Taylor, J. (1993). *Theories of International Migration: A Review and Appraisal.* Population and Development Review, 19 (3), pp. 431–466.

McCaskill, N. (2016). *Trump promises wall and massive deportation program*. POLITICO. Available at: https://www.politico.com/story/2016/08/donald-trump-immigration-address-arizona-227612 [Accessed: 22 April 2021].

Mechitishvili, N. (2021). *Rise of Populist Parties in the Era of Migration Crisis*. In: Abegunrin, O. and Abidde, S. (Eds.). *African Migrants and the Refugee Crisis*. Cham: Springer International Publishing, pp. 175–186.

Miguel, E. and Gugerty, M. (2005). *Ethnic diversity, social sanctions, and public goods in Kenya*. Journal of Public Economics, 89 (11–12), pp. 2325–2368.

National Institute of Statistics and Economic Studies (INSEE) (2021). *Definition - Purchasing power parity*. Available at: https://www.insee.fr/en/metadonnees/definition/c1923 [Accessed: 26 April 2021].

Organisation for Economic Co-operation and Development (OECD) (2020). *International Migration Outlook 2020*. Available at: http://www.oecd.org/migration/international-migration-outlook-1999124x.htm [Accessed: 6 March 2021].

Organisation for Economic Co-operation and Development (OECD) (2021a). *Land cover in countries and regions*. Available at: https://stats.oecd.org/Index.aspx?DataSetCode=LAND_COVER [Accessed: 21 April 2021].

Organisation for Economic Co-operation and Development (OECD) (2021b). *OECD Glossary of Statistical Terms*. Available at: https://stats.oecd.org/glossary/index.htm [Accessed: 25 March 2021].

Organisation for Economic Co-operation and Development (OECD) (2017) *Interrelations between Public Policies, Migration and Development*. Available at: https://read.oecd-ilibrary.org/development/interrelations-between-public-policies-migration-and-development_9789264265615-en [Accessed: 18 February 2021].

Ortega, F. and Peri, G. (2013). *The effect of income and immigration policies on international migration.* Migration Studies, 1 (1), pp. 47–74.

Rodrigue, J.-P. (2020). *The Geography of Transport Systems*. 5th ed. New York: Routledge.

Schewel, K. and Fransen, S. (2018). *Formal Education and Migration Aspirations in Ethiopia.* Population and Development Review. 44 (3), pp. 555–587.

Schiff, M. (2006). *Brain Gain: Claims About Its Size And Impact On Welfare And Growth Are Greatly Exaggerated.* In: Özden, Ç. and Schiff, M. (Eds.). *International Migration, Remittances and The Brain Drain.* The World Bank And Palgrave Macmillan, pp. 201–226.

Schumann, S. (2012). *Repräsentative Umfrage: Praxisorientierte Einführung in empirische Methoden und statistische Analyseverfahren*. 6th ed. München: De Gruyter Oldenbourg.

Schwab, K. (2017). *The Global Competitiveness Report 2017-2018.* Geneva: World Economic Forum.

Schwab, K. (2018). *The Global Competitiveness Report 2018*. Geneva: World Economic Forum.

Shilpi, F., Sangraula, P. and Li, Y. (2014). *Voting with Their Feet? Access to Infrastructure and Migration in Nepal*. The World Bank Policy Research Working Papers No 7047.

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

Stock, J. H. and Watson, M. W. (2006). *Introduction to Econometrics*. 2nd ed. Boston, Massachusetts: Addison Wesley.

United Nations (UN) (2021). *UNdata Glossary*. Available at: https://data.un.org/Glossary.aspx [Accessed: 25 March 2021].

United Nations Department of Economic and Social Affairs (UN DESA) (1998). *Recommendations on Statistics of International Migration*. Revision 1. New York: United Nations Statistical Papers Series M, No. 58.

United Nations Department of Economic and Social Affairs (UN DESA) (2018) *World Urbanization Prospects - Population Division - United Nations*. Available at: https://population.un.org/wup/ [Accessed: 23 April 2021].

United Nations Department of Economic and Social Affairs (UN DESA) (2021). *International Migrant Stock 2019 Documentation.* Available at: https://www.un.org/en/development/desa/population/migration/data/estimates2/docs/MigrationStockDocumentation_2019.pdf [Accessed: 23 April 2021].

United Nations Department of Economic and Social Affairs (UN DESA) (2021b). *World Population Prospects - Population Division - United Nations*. Available at: https://population.un.org/wpp/Download/Standard/Population/ [Accessed: 23 April 2021].

United Nations Department of Economic and Social Affairs (2021). *International Migrant Stock 2019*. Available at: https://www.un.org/en/development/desa/population/migration/data/estimates2/estimates19.asp [Accessed: 1 April 2021].

United Nations Food and Agriculture Organization (2021). *FAOSTAT*, *Value of Agricultural Production*. Available at: http://www.fao.org/faostat/en/#data/QV [Accessed: 3 April 2021].

United Nations Office on Drugs and Crime (2021). *Homicide rates*, *Tableau Software*. Available at: https://public.tableau.com/views/Homiciderates_15826327950430/Homicide-rates?:embed=y&:showVizHome=no&:host_url=https%3A%2F%2Fpublic.tableau.com%2F&:embed_code_version=3&:tabs=no&:toolbar=yes&:animate_transition=yes&:display_static_image=no&:display_spinner=no&:display_overlay=yes&:display_count=yes&:loadOrderID=0 [Accessed: 3 April 2021].

Webster, A. (2012). *Introductory Regression Analysis: With Computer Application for Business and Economics*. London: Taylor & Francis Group.

World Bank (2021a). *Global Bilateral Migration | DataBank*. Available at: https://databank.worldbank.org/reports.aspx?source=global-bilateral-migration [Accessed: 1 April 2021].

World Bank (2021b). *GNI per capita, PPP (constant 2017 international $) | Data*. Available at: https://data.worldbank.org/indicator/NY.GNP.PCAP.PP.KD [Accessed: 23 April 2021].

World Bank (2021c) *World Bank Country and Lending Groups – World Bank Data Help Desk*. Available at: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [Accessed: 28 April 2021].

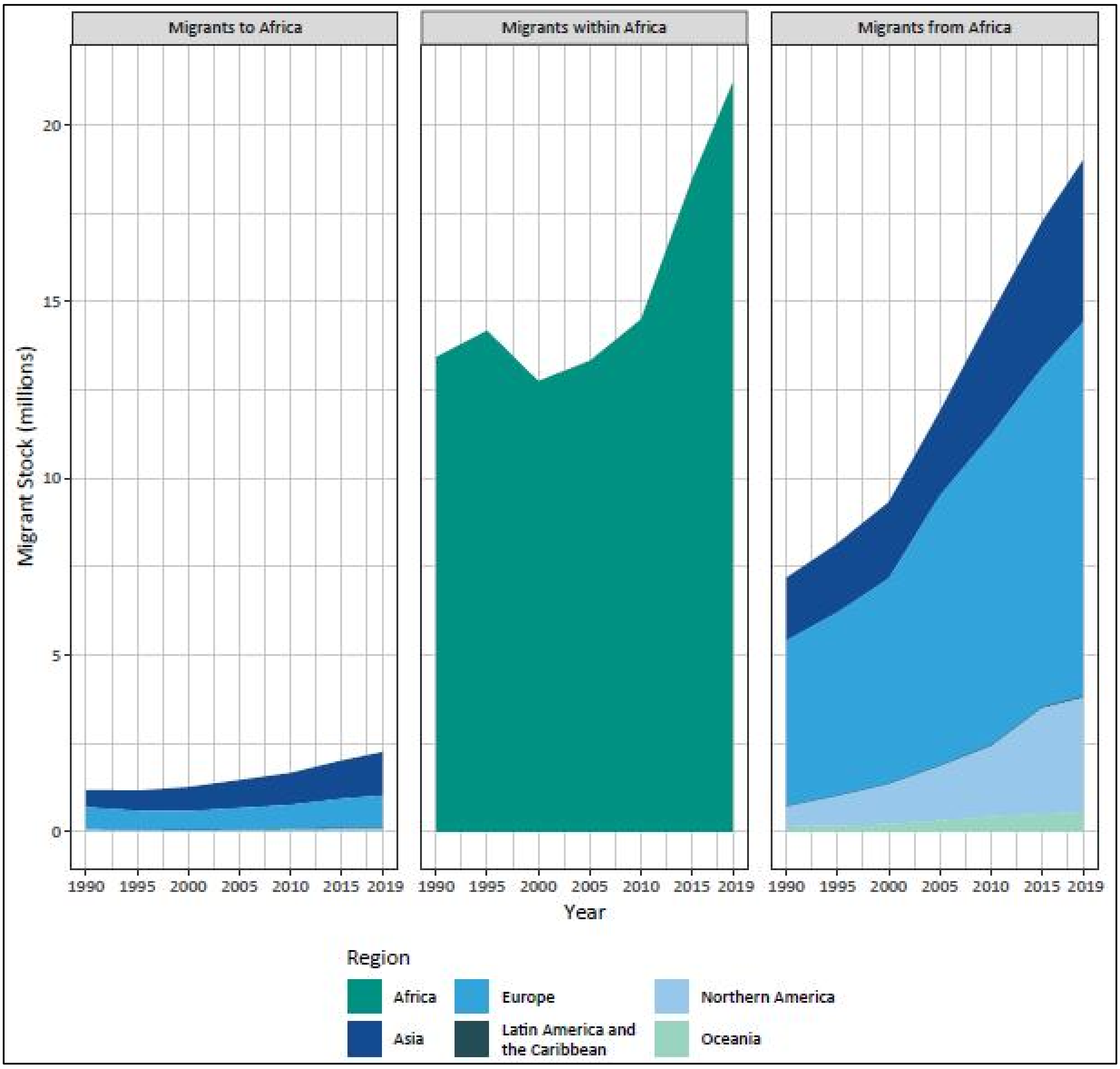
World Economic Forum (WEF) (2021). *Global Competitiveness Index 2017-2018*, *Global Competitiveness Index 2017-2018*. Available at: http://wef.ch/2wcj8jq [Accessed: 1 April 2021].

Xu, H., Lee, S. H. and Eom, T. (2008). Introduction to Panel Data Analysis. In: Yang, K. and Miller, G. J. (Eds.). *Handbook of Reserach Methods in Public Administration*.

Zelinsky, W. (1971). The Hypothesis of the Mobility Transition. *Geographical Review*, 61 (2), pp. 219–249.

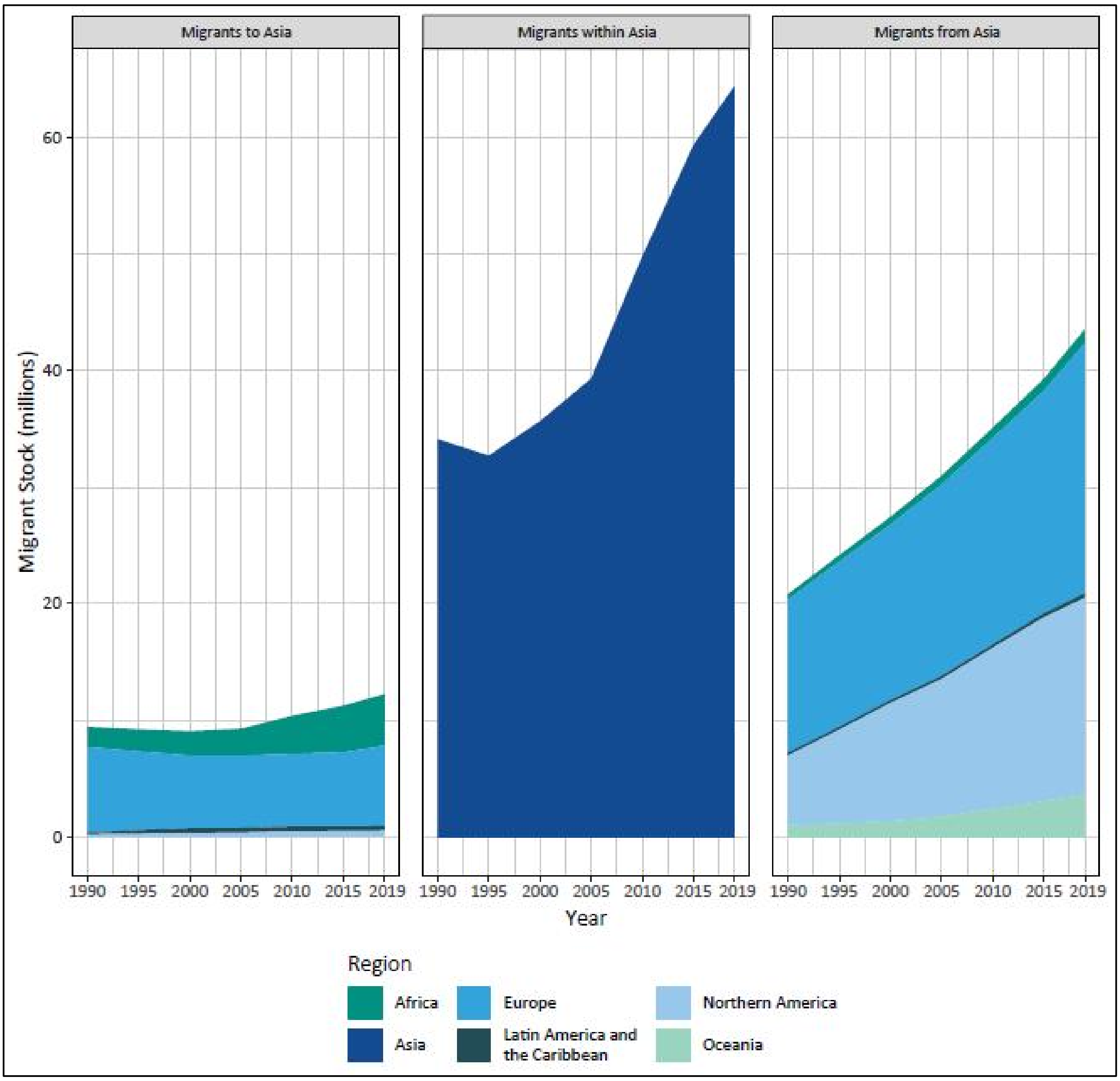
Appendix

Figure A-1: Migration trends in Africa



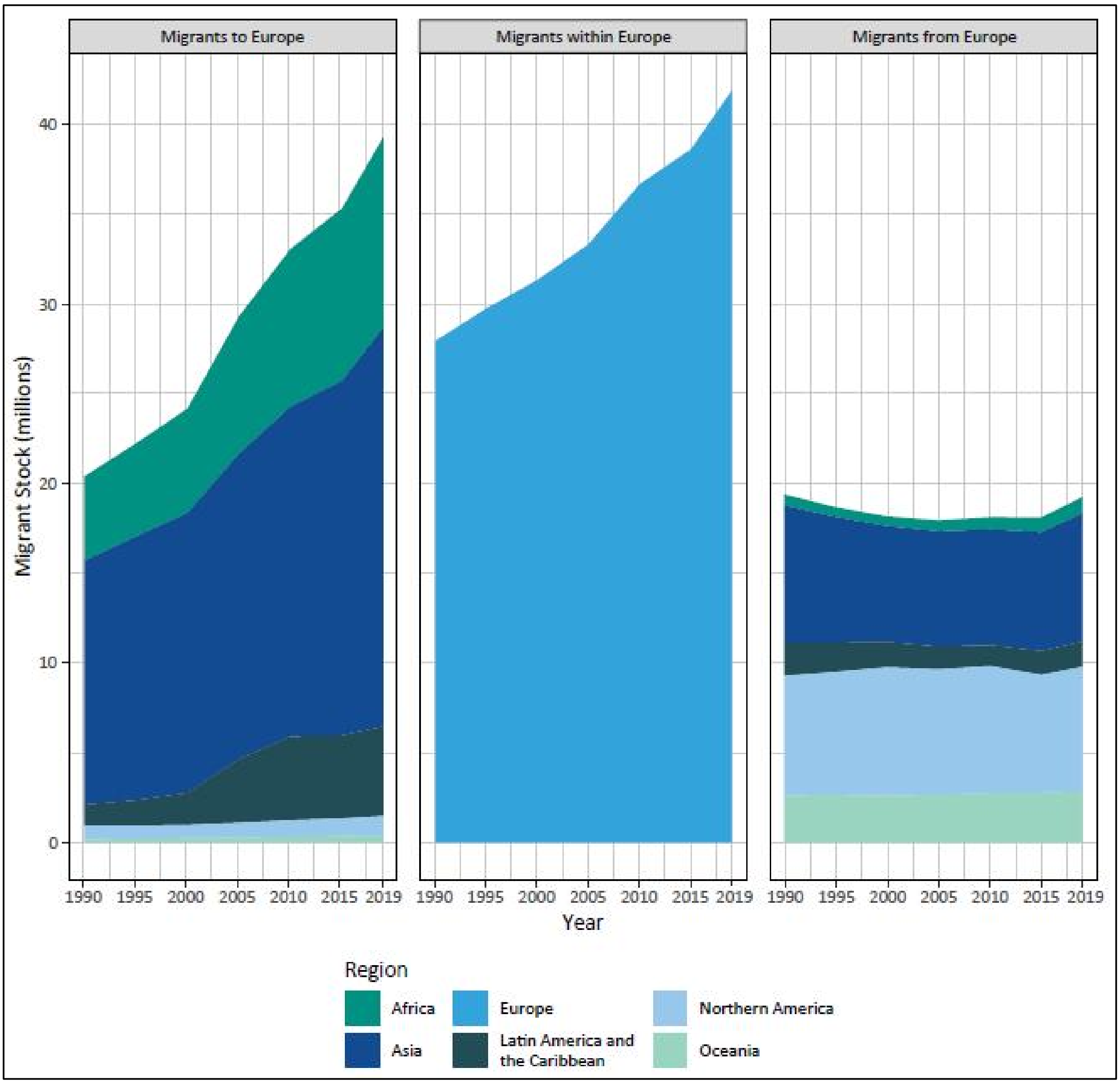
Source: IOM (2019a).

Figure A-2: Migration trends in Asia



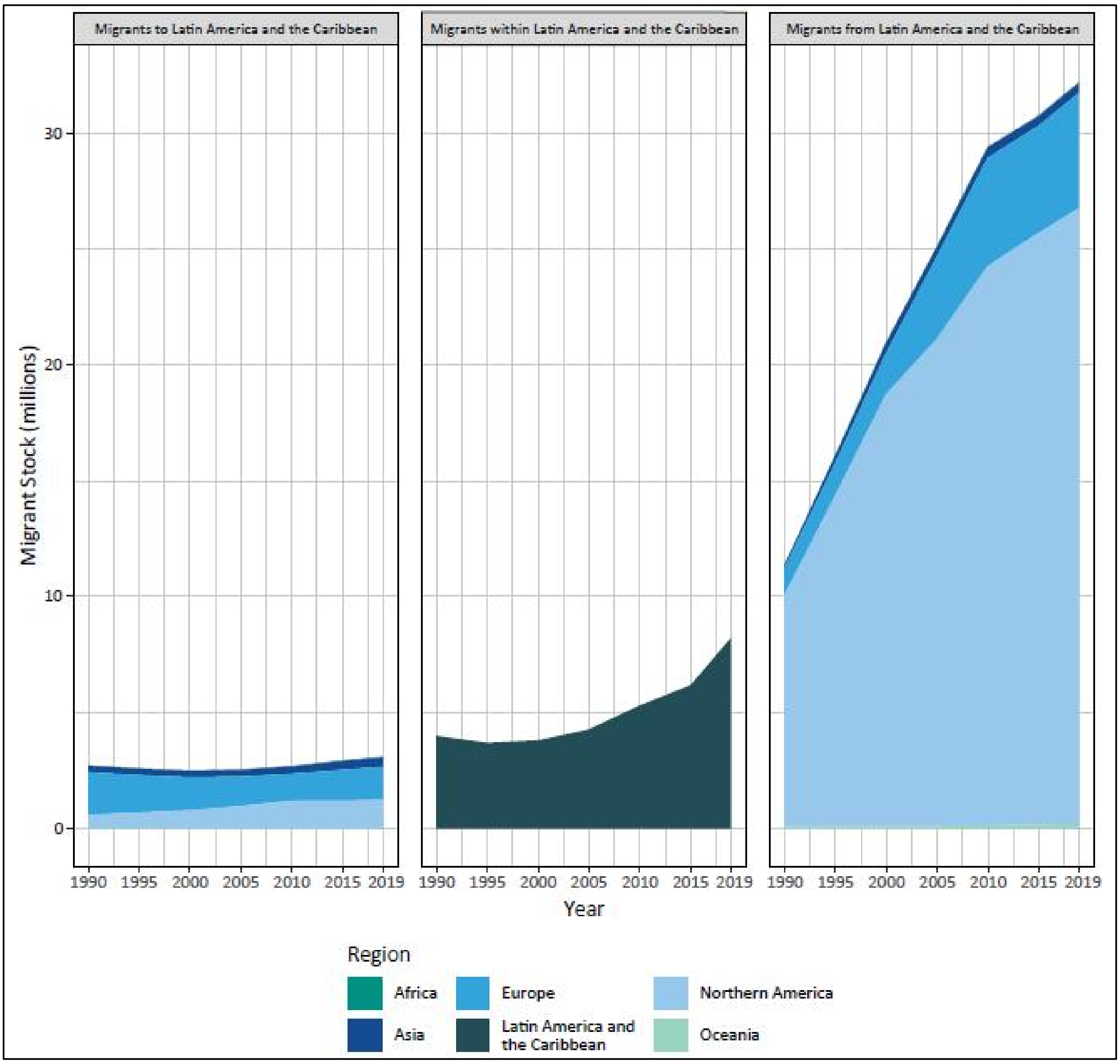
Source: IOM (2019a).

Figure A-3: Migration trends in Europe



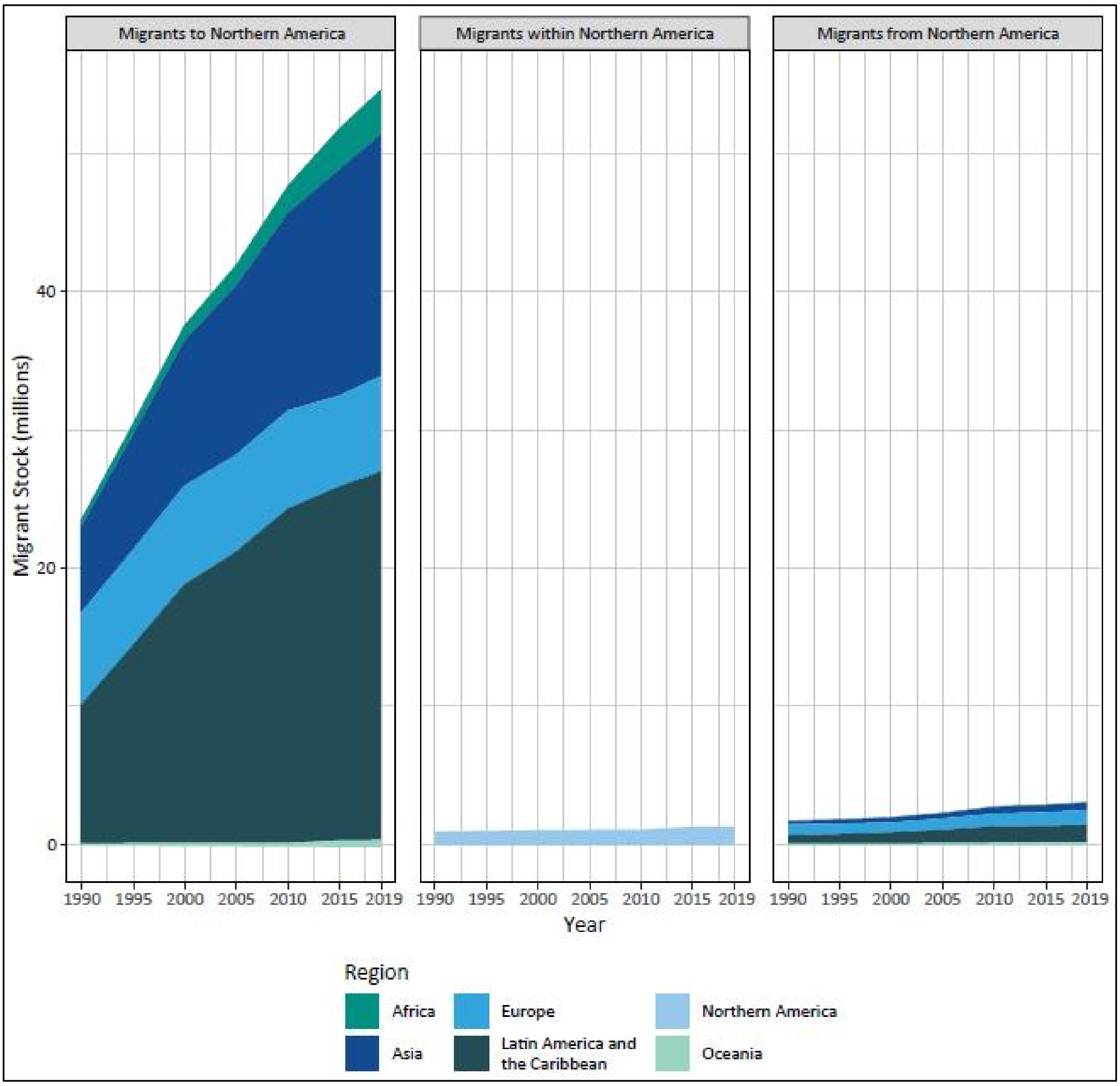
Source: IOM (2019a).

Figure A-4: Migration trends in Latin America and the Caribbean



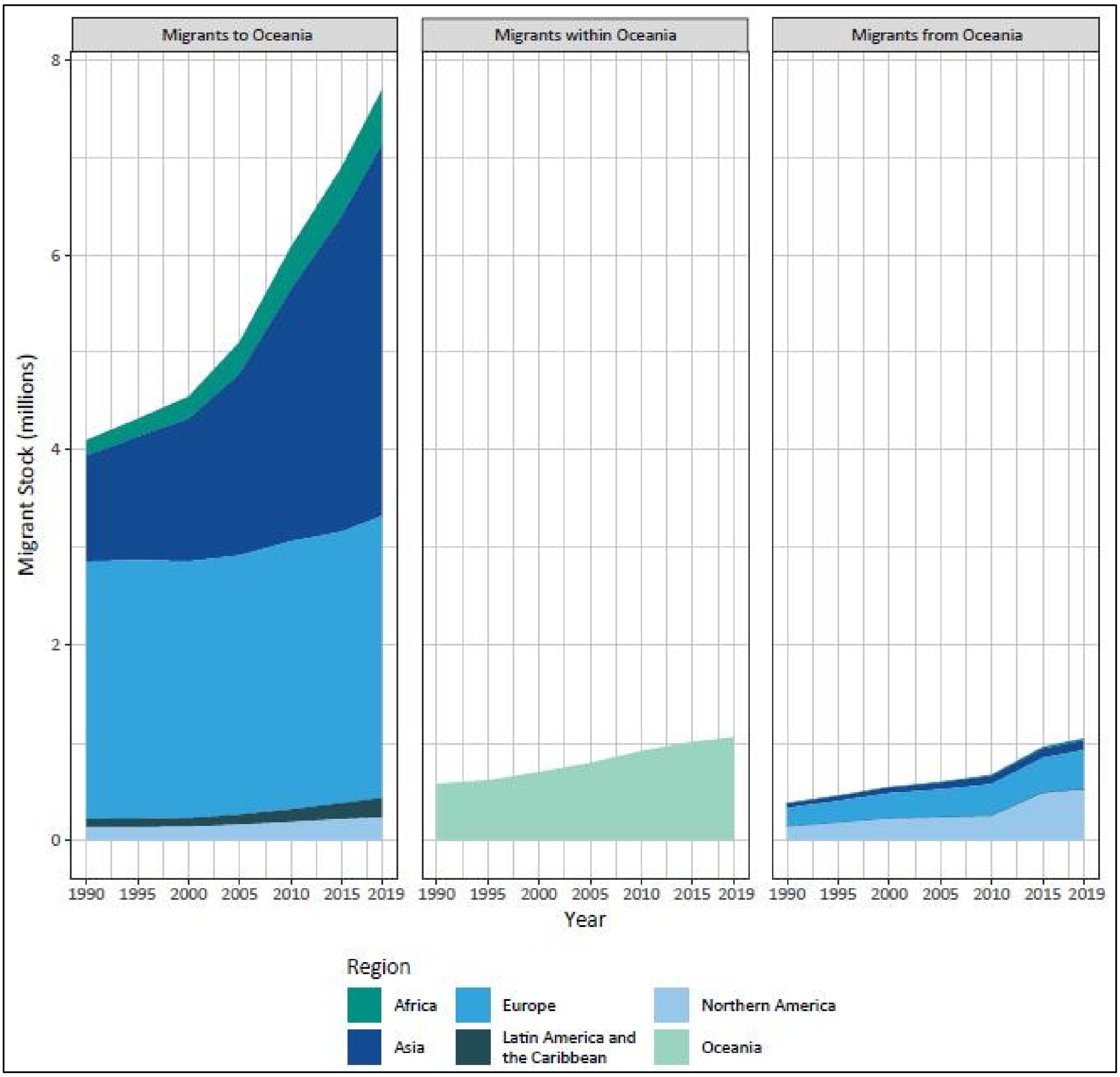
Source: IOM (2019a).

Figure A-5: Migration trends in Northern America



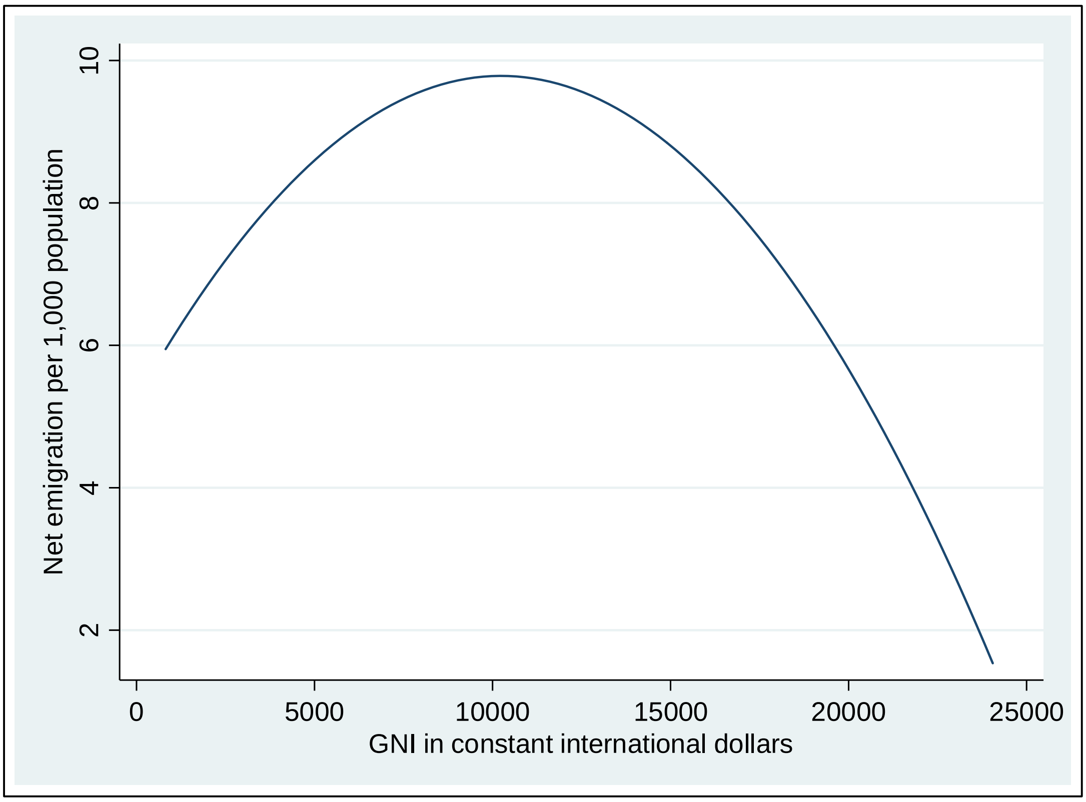
Source: IOM (2019a).

Figure A-6: Migration trends in Oceania



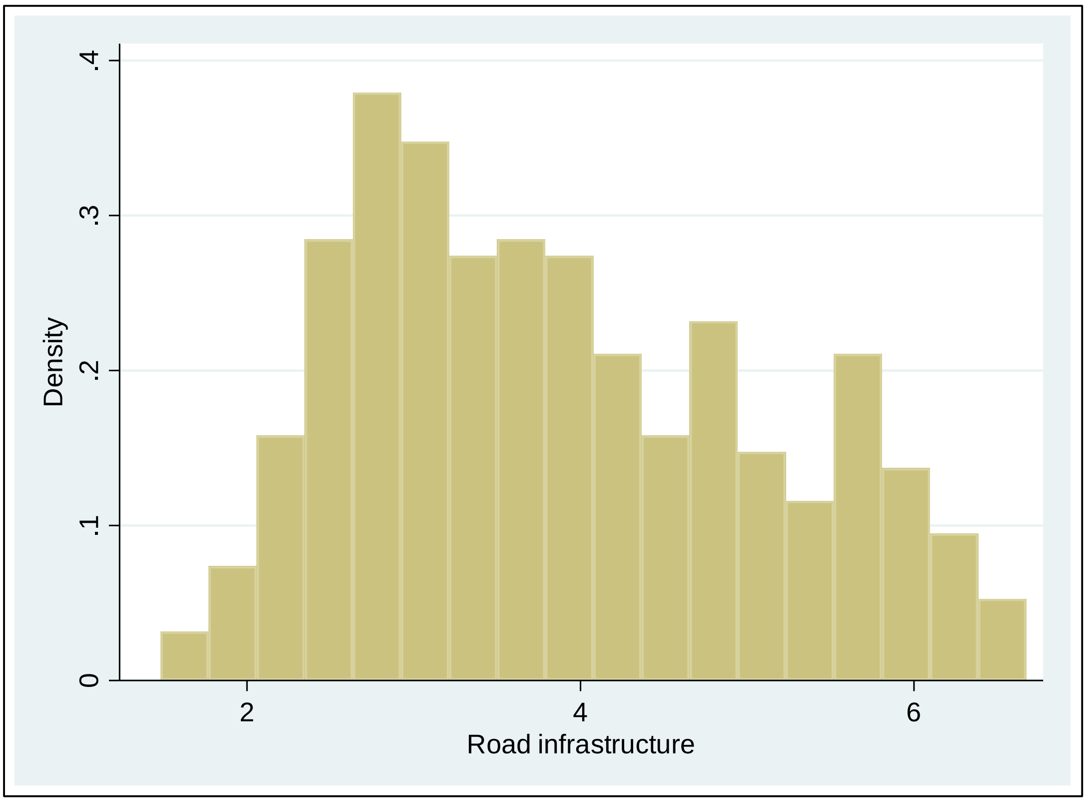
Source: IOM (2019a).

Figure A-7: Mobility-transition curve between 2005 and 2020



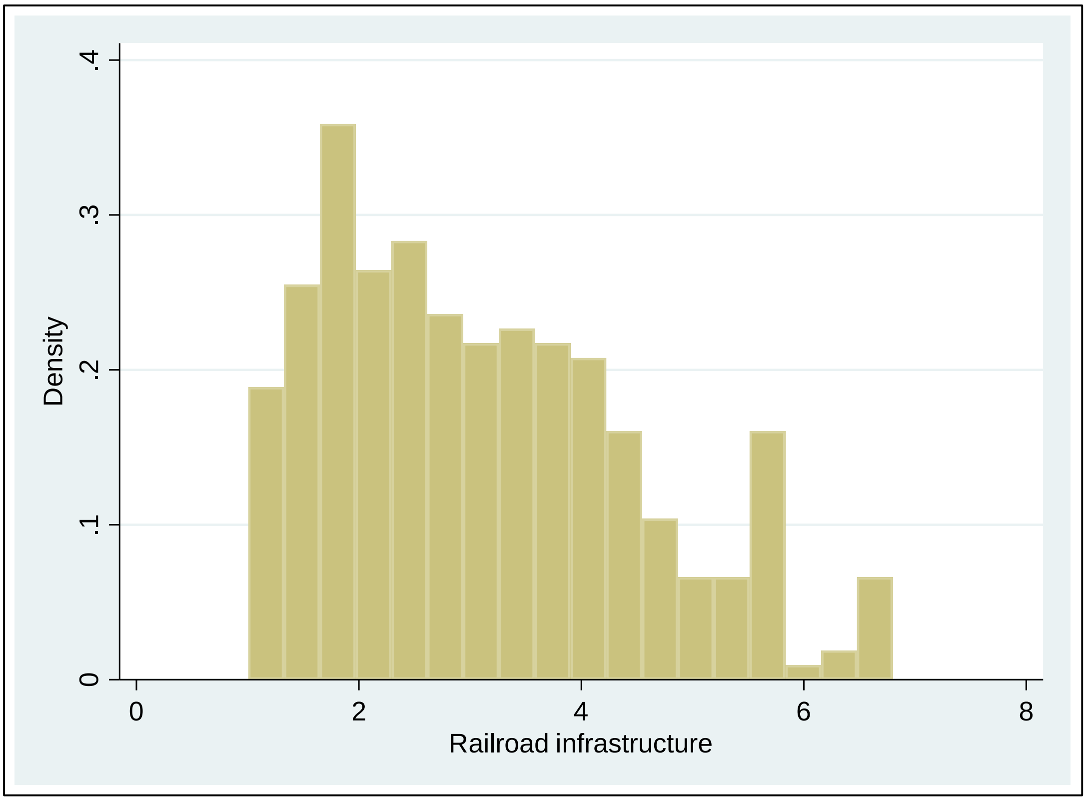
Source: Own compilation based on data from UN DESA (2021) and World Bank (2021b)
Note that the graph represents the fitted values for countries' average emigration rate in the stated period given the average value for GNI. Following (Clemens, 2020), countries that were classified in 2005 as high-income countries (World Bank classification) are excluded.

Figure A-8: Distribution of road infrastructure variable



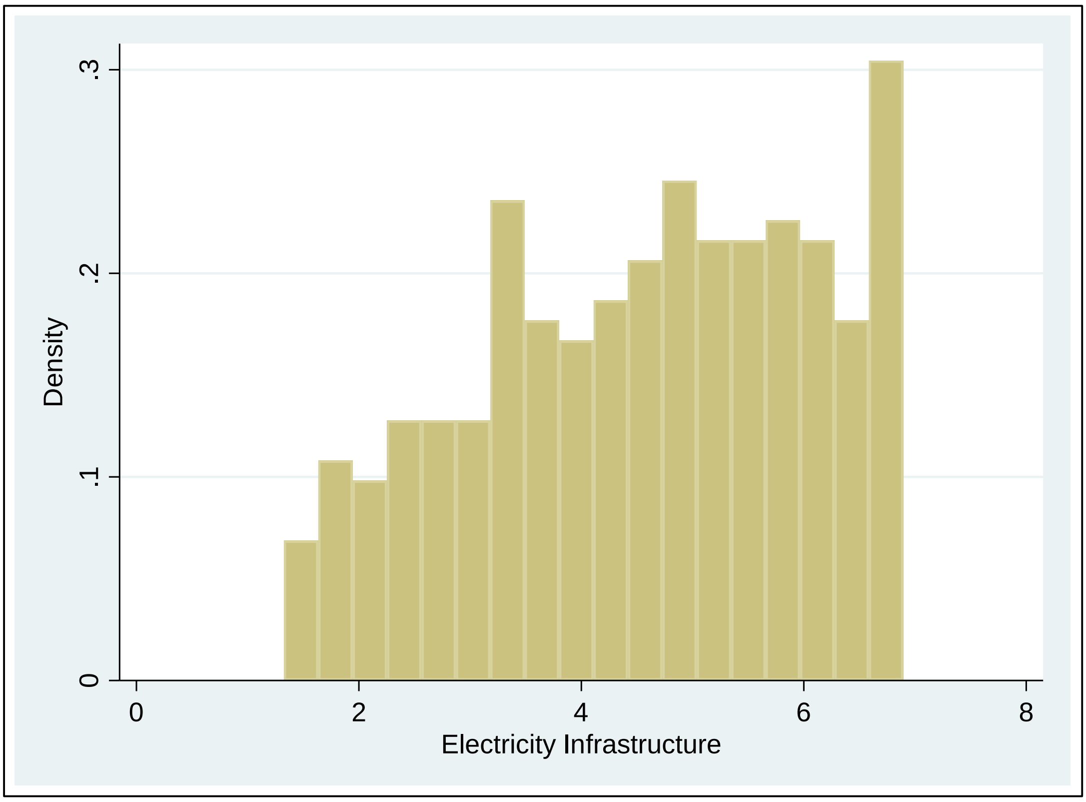
Source: Own compilation based on WEF (2021).

Figure A-9: Distribution of railroad infrastructure quality variable



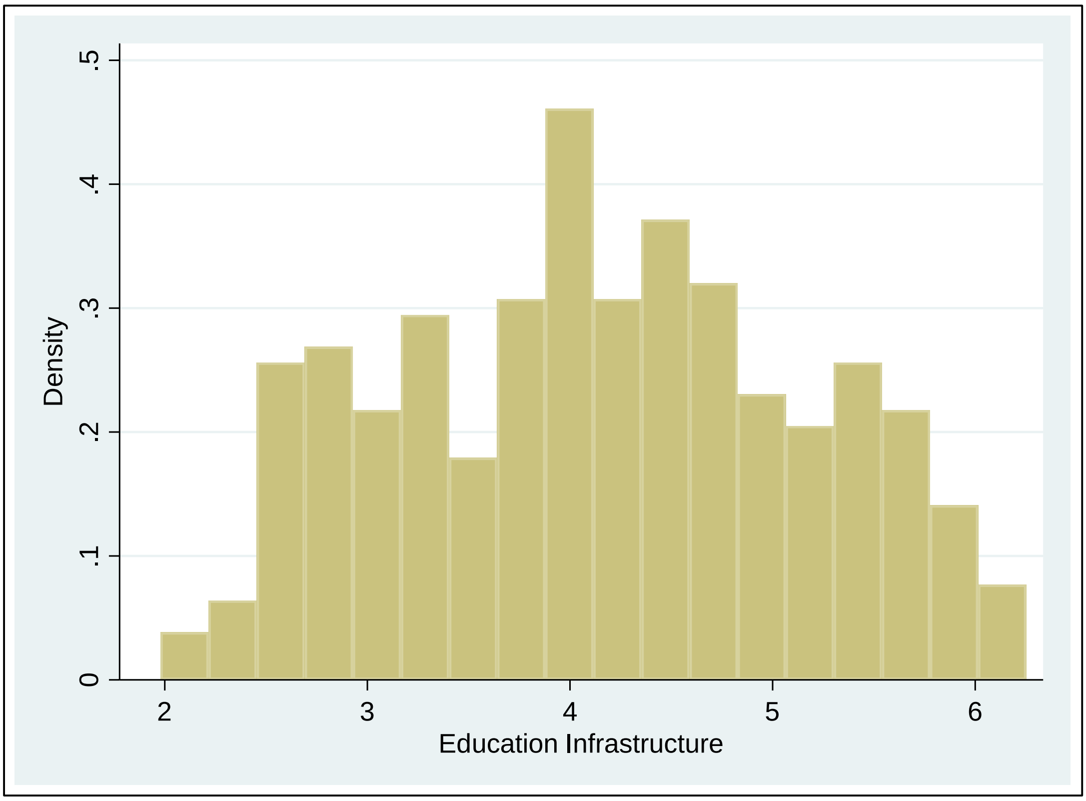
Source: Own compilation based on data from WEF (2021).

Figure A-10: Distribution of electricity infrastructure variable



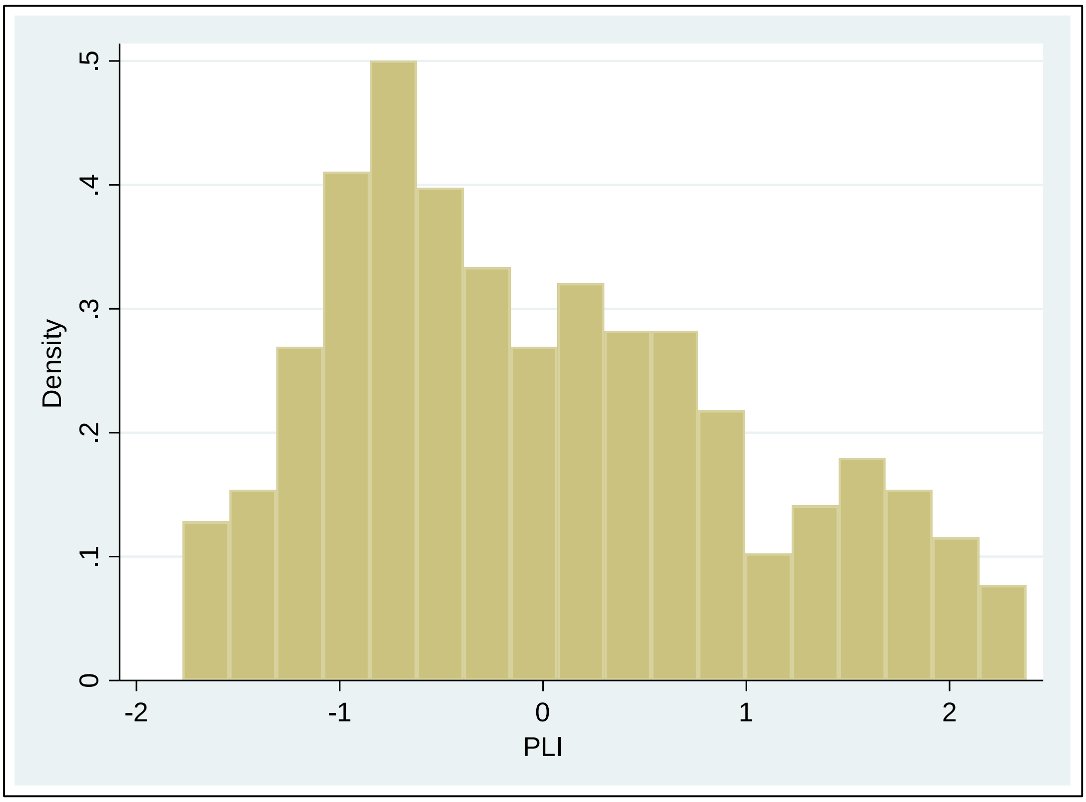
Source: Own compilation based on data from WEF (2021).

Figure A-11: Distribution of education infrastructure variable



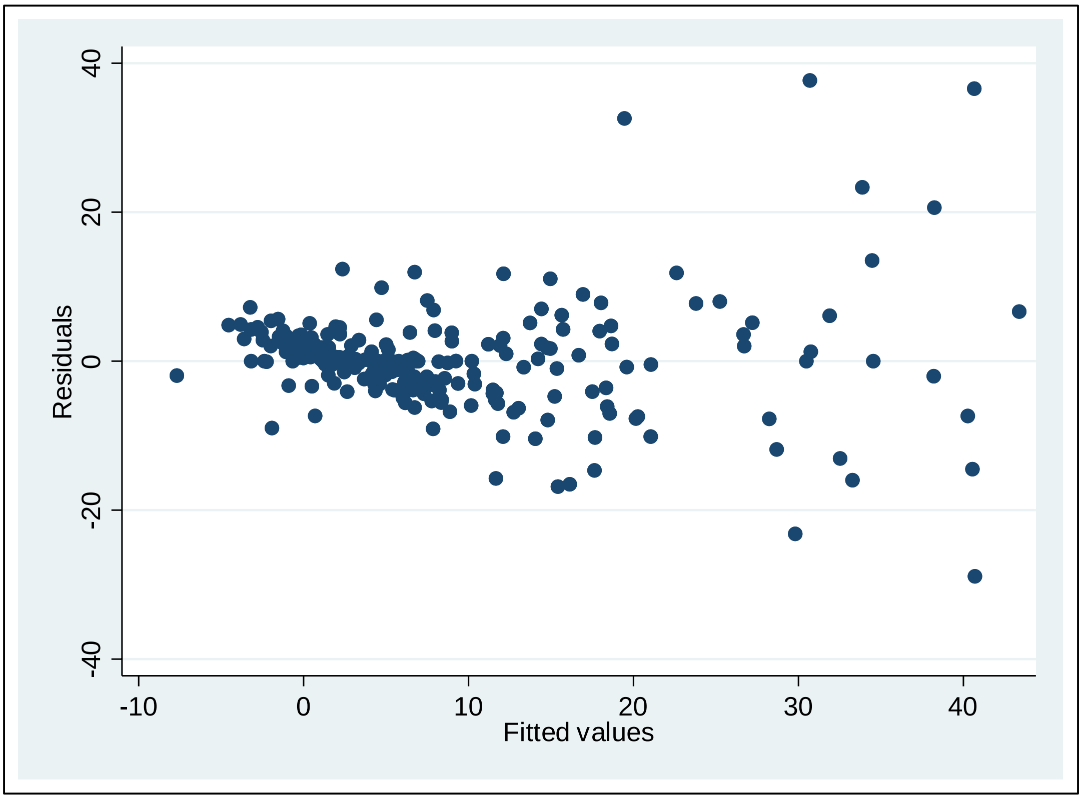
Source: Own compilation based on data from WEF (2021).

Figure A-12: Distribution of PLI index variable



Source: Own compilation based on data from WEF (2021).

Figure A-13: fe regression - distribution of residuals



Source: Own compilation based on fe regression model.

Table A-1: List of included countries

1. Albania	31. Ecuador	61. Lebanon	91. Portugal
2. Algeria	32. Egypt	62. Lesotho	92. Romania
3. Angola	33. El Salvador	63. Liberia	93. Russia
4. Argentina	34. Estonia	64. Libya	94. Saudi Arabia
5. Armenia	35. Eswatini	65. Lithuania	95. Senegal
6. Australia	36. Ethiopia	66. Madagascar	96. Serbia
7. Austria	37. Finland	67. Malawi	97. Sierra Leone
8. Azerbaijan	38. France	68. Malaysia	98. Singapore
9. Bangladesh	39. Georgia	69. Mali	99. Slovakia
10. Belgium	40. Germany	70. Mauritania	100. Slovenia
11. Benin	41. Ghana	71. Mexico	101. South Africa
12. Bolivia	42. Greece	72. Moldova	102. South Korea
13. Bosnia and Herzegovina	43. Guatemala	73. Mongolia	103. Spain
14. Botswana	44. Guinea	74. Morocco	104. Sri Lanka
15. Brazil	45. Haiti	75. Mozambique	105. Sweden
16. Bulgaria	46. Honduras	76. Myanmar	106. Switzerland
17. Burkina Faso	47. Hungary	77. Namibia	107. Tajikistan
18. Cambodia	48. India	78. Nepal	108. Tanzania
19. Cameroon	49. Indonesia	79. Netherlands	109. Thailand
20. Canada	50. Iran	80. New Zealand	110. Tunisia
21. Chile	51. Ireland	81. Nicaragua	111. Turkey
22. China	52. Israel	82. Nigeria	112. Uganda
23. Colombia	53. Italy	83. North Macedonia	113. Ukraine
24. Congo (Democratic Republic)	54. Jamaica	84. Norway	114. United Kingdom
25. Costa Rica	55. Japan	85. Pakistan	115. United States
26. Croatia	56. Jordan	86. Panama	116. Uruguay
27. Czechia	57. Kazakhstan	87. Paraguay	117. Venezuela
28. Côte d'Ivoire	58. Kenya	88. Peru	118. Vietnam
29. Denmark	59. Kyrgyzstan	89. Philippines	119. Zambia
30. Dominican Republic	60. Latvia	90. Poland	120. Zimbabwe

Table A-2: Survey questions of the Global Competitiveness Index

Road infrastructure	In your country, what is the quality (extensiveness and condition) of road infrastructure? [1 = extremely poor—among the worst in the world; 7 = extremely good—among the best in the world]
Railroad infrastructure	In your country, what is the quality (extensiveness and condition) of the railroad system? [1 = extremely poor—among the worst in the world; 7 = extremely good—among the best in the world]
Electricity infrastructure	In your country, how reliable is the electricity supply (lack of interruptions and lack of voltage fluctuations)? [1 = extremely unreliable; 7 = extremely reliable]
Education infrastructure	In your country, how well does the education system meet the needs of a competitive economy? [1 = extremely unreliable; 7 = extremely reliable]

Source: Based on WEF (2021).

Table A-3: Regression table – fe with interaction term

	(1)	(2)	(3)	(4)
	Emigration	Emigration	Emigration	Emigration
PLI	-15.24	-29.50*	-17.10	-28.85**
	(12.85)	(16.76)	(15.69)	(14.49)
PEI	7.488	9.948	13.01	8.764
	(6.114)	(8.077)	(8.925)	(8.322)
PLI*PEI	2.969	5.484	2.799	6.319**
	(3.496)	(4.067)	(3.865)	(3.169)
Conflict		-0.202	-0.318	-0.152
		(0.351)	(0.415)	(0.199)
Freedom		3.519	2.801	-0.819
		(5.727)	(5.578)	(3.236)
Homicides		-0.0512	-0.0496	0.0804
		(0.363)	(0.396)	(0.322)
Natural Disasters		-0.0239	-0.00898	-0.0192
		(0.0358)	(0.0341)	(0.0295)
Median Age			0.337	0.0546
			(1.748)	(1.794)
Population Density			-0.164	-0.117
			(0.121)	(0.114)
Urbanization			-0.283	0.0253
			(0.773)	(0.721)
Migrant Stock			-0.399***	-0.342***
			(0.118)	(0.127)
GNI				-0.219
				(0.674)
Observations	329	258	252	237
R-squared	0.037	0.077	0.238	0.235
Number of Countries	120	105	103	99

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table A-4: Regression table – fe without high-income countries

	(1)	(2)	(3)	(4)	(5)
	Emigration	Emigration	Emigration	Emigration	Emigration
PLI	-6.255	-13.47	-9.982	-6.713	-61.53*
	(8.150)	(10.23)	(9.109)	(7.635)	(32.86)
PEI	7.780	9.370	11.37	0.745	9.970
	(7.919)	(8.510)	(8.430)	(8.010)	(9.637)
PLI*PEI					15.04*
					(8.826)
Conflict		-0.309	-0.242	-0.0374	0.100
		(0.370)	(0.433)	(0.315)	(0.351)
Freedom		3.057	2.931	-2.541	-2.010
		(5.895)	(6.236)	(4.347)	(4.481)
Homicides		0.00996	0.0214	0.284	0.146
		(0.398)	(0.440)	(0.318)	(0.327)
Natural Disasters		-0.0237	-0.0125	-0.0158	-0.0255
		(0.0427)	(0.0411)	(0.0359)	(0.0393)
Median Age			0.403	-1.284	-0.794
			(2.241)	(2.931)	(2.944)
Population Density			-0.103	-0.0475	0.0119
			(0.106)	(0.0886)	(0.0991)
Urbanization			-0.743	-0.146	-0.232
			(1.101)	(1.031)	(1.082)
Migrant Stock			-0.264	-0.101	-0.0823
			(0.210)	(0.252)	(0.243)
GNI				2.316	0.530
				(2.321)	(2.424)
Observations	215	164	161	148	148
R-squared	0.056	0.138	0.205	0.210	0.256
Number of Countries	82	71	70	66	66

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table A-5: Regression table – re with interaction term

	(1)	(2)	(3)	(4)
	Emigration	Emigration	Emigration	Emigration
PLI	-0.440	-3.288	-5.864	-4.759
	(3.947)	(4.583)	(4.453)	(4.273)
PEI	2.583	1.951	-1.155	-3.068*
	(1.828)	(2.190)	(2.416)	(1.859)
PLI*PEI	-0.764	-0.0405	0.721	0.954
	(0.758)	(0.868)	(0.802)	(0.892)
Conflict		-0.201	-0.105	0.103
		(0.390)	(0.366)	(0.203)
Freedom		-0.960	-0.654	-1.082**
		(0.696)	(0.807)	(0.489)
Homicides		0.188*	0.212**	0.111**
		(0.0967)	(0.0972)	(0.0552)
Natural Disaster		0.0293	0.0361*	0.0418**
		(0.0220)	(0.0191)	(0.0186)
Median Age			0.345*	0.351*
			(0.208)	(0.193)
Population Disaster			0.000151	0.000609
			(0.000519)	(0.000614)
Urbanization			-0.00357	0.0286
			(0.0577)	(0.0626)
Migrant Stock			0.0536***	0.0677***
			(0.0161)	(0.0145)
GNI				-0.106
				(0.108)
Constant	-1.957	1.481	-1.538	7.034
	(7.334)	(10.72)	(12.27)	(7.569)
Observations	329	258	252	237
Overall R-squared	0.040	0.082	0.210	0.274
Number of Countries	120	105	103	99

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A-6: Regression table – pooled OLS

	(1)	(2)	(3)	(4)	(5)
	Emigration	Emigration	Emigration	Emigration	Emigration
PLI	-3.826**	-3.370*	-2.461	-0.657	-4.759
	(1.628)	(1.758)	(1.863)	(1.535)	(4.273)
PEI	2.700*	1.696	-1.489	-3.720**	-3.068
	(1.526)	(1.945)	(2.305)	(1.692)	(1.859)
PLI*PEI					0.954
					(0.892)
Conflict		-0.177	-0.115	0.0791	0.103
		(0.378)	(0.361)	(0.201)	(0.203)
Freedom		-1.105*	-0.775	-1.137**	-1.082**
		(0.622)	(0.765)	(0.464)	(0.489)
Homicides		0.180**	0.201**	0.104*	0.111**
		(0.0882)	(0.0934)	(0.0550)	(0.0552)
Natural Disasters		0.0354	0.0372**	0.0406**	0.0418**
		(0.0225)	(0.0189)	(0.0184)	(0.0186)
Median Age			0.338*	0.332*	0.351*
			(0.203)	(0.195)	(0.193)
Population Density			0.000476	0.000801	0.000609
			(0.000444)	(0.000560)	(0.000614)
Urbanization			-0.00196	0.0197	0.0286
			(0.0563)	(0.0608)	(0.0626)
Migrant Stock			0.0535***	0.0669***	0.0677***
			(0.0156)	(0.0145)	(0.0145)
GNI				-0.0478	-0.106
				(0.0920)	(0.108)
Constant	-3.291	2.868	0.892	10.70*	7.034
	(6.135)	(9.293)	(11.29)	(6.183)	(7.569)
Observations	329	258	252	237	237
Adjusted R-squared	0.028	0.06	0.175	0.235	0.235

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table A-7: Regression table – fe with separate PLI

	(1)	(2)	(3)	(4)
	Emigration	Emigration	Emigration	Emigration
Road	1.212	1.850	1.901	2.190
	(2.719)	(3.977)	(4.314)	(4.520)
Railroad	-1.764	-3.490	-6.302	-5.445
	(4.464)	(6.731)	(5.665)	(5.658)
Electricity	-3.611	-6.204	-3.185	-2.063
	(3.375)	(4.100)	(3.595)	(4.060)
Education	5.549	7.433	13.41	9.533
	(5.147)	(4.982)	(8.787)	(8.477)
Conflict		-0.0720	-0.210	-0.0571
		(0.382)	(0.498)	(0.277)
Freedom		3.050	3.372	-0.410
		(4.701)	(5.204)	(3.158)
Homicides		-0.0910	-0.131	0.0716
		(0.412)	(0.441)	(0.356)
Natural Disasters		-0.0269	-0.0136	-0.0224
		(0.0279)	(0.0265)	(0.0232)
Median Age			1.392	1.115
			(1.250)	(1.632)
Population Density			-0.149	-0.130
			(0.103)	(0.095)
Urbanization			-0.397	-0.158
			(0.754)	(0.726)
Migrant Stock			-0.420***	-0.400***
			(0.117)	(0.130)
GNI				0.284
				(0.707)
Constant	2.017	1.877	6.919	8.456
	(17.50)	(25.02)	(27.49)	(28.74)
Observations	329	258	252	237
Within R-squared	0.022	0.058	0.227	0.202
Number of Countries	120	105	103	99

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table A-8: Regression table – re with separate PLI

	(1)	(2)	(3)	(4)
	Emigration	Emigration	Emigration	Emigration
Road	1.121	0.678	1.615	1.109
	(1.173)	(1.258)	(1.255)	(1.293)
Railroad	-3.922***	-2.859**	-2.551**	-1.497
	(1.150)	(1.252)	(1.235)	(1.105)
Electricity	-0.319	-0.772	-1.875	-0.412
	(1.380)	(1.658)	(1.750)	(1.158)
Education	2.832	2.411	-0.600	-3.345**
	(2.143)	(2.750)	(2.603)	(1.704)
Conflict		-0.177	-0.112	0.0755
		(0.408)	(0.381)	(0.212)
Freedom		-0.745	-0.534	-0.968**
		(0.604)	(0.719)	(0.437)
Homicides		0.137	0.168*	0.0826
		(0.0977)	(0.0948)	(0.0648)
Natural Disasters		0.0264	0.0338*	0.0370*
		(0.0223)	(0.0194)	(0.0195)
Median Age			0.468**	0.410**
			(0.227)	(0.207)
Population Density			0.0003	0.0008
			(0.0005)	(0.0006)
Urbanization			-0.0107	0.0107
			(0.0568)	(0.0560)
Migrant Stock			0.0494***	0.0637***
			(0.0154)	(0.0141)
GNI				-0.0568
				(0.0950)
Constant	5.573*	9.029	3.802	9.603
	(3.233)	(7.408)	(8.343)	(6.142)
Observations	329	258	252	237
Overall R-squared	0.081	0.097	0.225	0.278
Number of Countries	120	105	103	99

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table A-9: Regression table – pooled OLS with separate PLI

	(1)	(2)	(3)	(4)
	Emigration	Emigration	Emigration	Emigration
Road	0.930	0.571	1.602	1.109
	(1.002)	(1.180)	(1.222)	(1.293)
Railroad	-3.932***	-2.818**	-2.515**	-1.497
	(0.940)	(1.145)	(1.198)	(1.105)
Electricity	0.261	-0.514	-1.846	-0.412
	(1.184)	(1.550)	(1.713)	(1.158)
Education	2.328	2.036	-0.725	-3.345*
	(1.812)	(2.535)	(2.554)	(1.704)
Conflict		-0.168	-0.104	0.0755
		(0.402)	(0.379)	(0.212)
Freedom		-0.827	-0.558	-0.968**
		(0.563)	(0.703)	(0.437)
Homicides		0.131	0.166*	0.0826
		(0.0909)	(0.0927)	(0.0648)
Natural Disaster		0.0308	0.0356*	0.0370*
		(0.0228)	(0.0194)	(0.0195)
Median Age			0.467**	0.410**
			(0.222)	(0.207)
Population Density			0.0003	0.0008
			(0.0005)	(0.0006)
Urbanization			-0.00738	0.0107
			(0.0557)	(0.0560)
Migrant Stock			0.0505***	0.0637***
			(0.0149)	(0.0141)
GNI				-0.0568
				(0.0950)
Constant	5.651**	9.916	3.948	9.603
	(2.616)	(6.890)	(8.154)	(6.142)
Observations	329	258	252	237
Adjusted R-squared	0.071	0.069	0.186	0.236

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table A-10: Regression table – cross-country with separate PLI

	(1)	(2)	(3)	(4)
	Emigration	Emigration	Emigration	Emigration
Road	0.690	0.0647	1.375	1.382
	(1.511)	(1.678)	(1.188)	(1.226)
Railroad	-5.828***	-4.578***	-3.844**	-3.745**
	(1.352)	(1.588)	(1.489)	(1.475)
Electricity	1.386	1.451	-0.462	0.873
	(1.648)	(1.911)	(1.792)	(1.682)
Education	2.255	1.586	-0.605	-1.131
	(1.661)	(1.811)	(2.295)	(2.178)
Freedom		-0.435	-0.422	-0.812
		(0.571)	(0.661)	(0.621)
Conflict		-0.0895	-0.0265	0.0168
		(0.532)	(0.385)	(0.328)
Homicides		0.146*	0.138*	0.0677
		(0.0786)	(0.0790)	(0.0668)
Natural Disasters		$-2.68*10^{-8}$	$2.58*10^{-8}$	$3.22*10^{-8}$
		$(3.97*10^{-8})$	$(4.31*10^{-8})$	$(4.54*10^{-8})$
Median Age			0.305	0.233
			(0.213)	(0.213)
Population Density			0.0009**	0.0013***
			(0.0004)	(0.0004)
Urbanization			-0.0255	-0.0391
			(0.0564)	(0.0577)
Migrant Stock			0.0609***	0.0650***
			(0.0147)	(0.0145)
GNI				-0.0416
				(0.0918)
Intercept	1.152	3.704	-2.040	4.351
	(5.071)	(6.412)	(8.664)	(7.796)
Observations	120	106	104	103
Adjusted R-squared	0.14	0.142	0.421	0.430

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table A-11: Regression table – cross-country with interaction term

	(1)	(2)	(3)	(4)
	Emigration	Emigration	Emigration	Emigration
PLI	1.782	0.936	-1.613	-0.733
	(3.860)	(4.339)	(4.371)	(4.452)
PEI	7.804***	5.531*	-0.235	-0.592
	(2.591)	(3.173)	(2.822)	(2.922)
PLI*PEI	-1.700**	-1.229	-0.191	-0.196
	(0.824)	(1.010)	(0.905)	(1.026)
Freedom		-0.752	-0.640	-0.988
		(0.627)	(0.728)	(0.709)
Conflict		-0.185	0.0381	0.0761
		(0.528)	(0.394)	(0.349)
Homicides		0.192**	0.174**	0.118*
		(0.0765)	(0.0747)	(0.0659)
Natural Disasters		$-8.88*10^{-8}$	$1.41*10^{-9}$	$7.98*10^{-9}$
		$(5.77*10^{-8})$	$4.54*10^{-8}$	$5.01*10^{-8}$
Median Age			0.198	0.147
			(0.209)	(0.210)
Population Density			0.00095**	0.0012**
			(0.0004)	(0.0005)
Urbanization			-0.00365	-0.0124
			(0.0584)	(0.0624)
Migrant Stock			0.0656***	0.0703***
			(0.0147)	(0.0143)
GNI				-0.0235
				(0.117)
Constant	-4.204	1.826	5.529	7.377
	(8.443)	(10.31)	(9.730)	(9.682)
Observations	120	106	104	103
Adjusted R-squared	0.071	0.108	0.394	0.394

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Table A-12: Regression table – instrumental variable without high-income countries

	(1)	(2)
	PLI	Emigration
Wetland	-0.0746***	
	(0.0235)	
Freedom	0.0374	-0.210
	(0.0458)	(0.905)
Homicides	0.00486	0.221***
	(0.00429)	(0.0793)
Natural Disasters	$4.51*10^{-8}$**	$9.54*10^{-7}$***
	$1.72*10^{-8}$	$3.42*10^{-7}$
Conflict	0.0183	0.240
	(0.0315)	(0.510)
Median Age	0.00986	0.263
	(0.0171)	(0.327)
Population Density	-0.0006	-0.0067*
	(0.0004)	(0.0041)
Urbanization	-0.00543	-0.112
	(0.00619)	(0.128)
Migrant Stock	-0.000850	0.0482***
	(0.000624)	(0.0153)
GNI	0.0621***	0.731
	(0.0221)	(0.535)
Agricultural Production	$-2.86*10^{-9}$*	$-7.60*10^{-8}$***
	(1.57e-09)	(2.95e-08)
PLI		-11.66**
		(5.141)
Constant	2.092***	27.16**
	(0.437)	(12.17)
Observations	70	70
R-squared	0.380	0.126

Robust standard errors in parentheses
**** p<0.01, ** p<0.05, * p<0.1*

Public Policy Master’s Thesis Series

This series presents the Master’s theses in Public Policy and in European Affairs of the Sciences Po School of Public Affairs. It aims to promote high-standard research master’s theses, relying on interdisciplinary analyses and leading to evidence-based policy recommendations.

Benefits on-site or to take away:

A panel data analysis on how infrastructures shape emigration

Tobias Hillenbrand

Abstract

In light of upward-pointing migration trends, policymakers continue to suggest socioeconomic development as a means to reduce the number of international immigrants despite plenty of contradicting evidence. However, uncertainty remains about what exactly drives the emigration rates of lower-income countries on their path to economic prosperity. As the role of infrastructures is underexplored, I contribute to the investigation of the infrastructure-emigration relationship by proposing a novel classification that distinguishes between an infrastructure type with benefits tied to locations (place-based) and another type with benefits tied to persons (people-based). Due to differences in the “portability” of benefits, I hypothesize that the place-based type affects emigration negatively, whereas people-based infrastructures should have a positive impact. I compile a panel dataset with data from UN DESA, the Global Competitiveness Index, and other sources, that includes 120 countries and covers a period from 2005 to 2020. My hypotheses are tested through the application of different regression analysis techniques, among them are fixed effects and instrumental variable regressions. The findings suggest that place-based infrastructures are likely to affect emigration negatively, while the picture for education – representing the people-based type – is more mixed. The results are of high political relevance and encourage further research (e. g., using individual-level data) to create a deeper understanding of the patterns of cross-border movements.

Key words

migration, emigration, infrastructures, place-based, people-based, development