European University Institute

ROBERT SCHUMAN CENTRE FOR ADVANCED STUDIES

EUI WORKING PAPERS

RSCAS 2019/30
Robert Schuman Centre for Advanced Studies
Migration Policy Centre

Estimating Transnational Human Mobility on a Global Scale

Ettore Recchi, Emanuel Deutschmann and Michele Vespe

European University Institute

Robert Schuman Centre for Advanced Studies

Migration Policy Centre

Estimating Transnational Human Mobility on a Global Scale

Ettore Recchi, Emanuel Deutschmann and Michele Vespe

EUI Working Paper **RSCAS** 2019/30

ISSN 1028-3625

Printed in Italy, April 2019
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu/RSCAS/Publications/
www.eui.eu
cadmus.eui.eu

Robert Schuman Centre for Advanced Studies

The Robert Schuman Centre for Advanced Studies, created in 1992 and currently directed by Professor Brigid Laffan, aims to develop inter-disciplinary and comparative research on the major issues facing the process of European integration, European societies and Europe's place in 21st century global politics.

The Centre is home to a large post-doctoral programme and hosts major research programmes, projects and data sets, in addition to a range of working groups and *ad hoc* initiatives. The research agenda is organised around a set of core themes and is continuously evolving, reflecting the changing agenda of European integration, the expanding membership of the European Union, developments in Europe's neighbourhood and the wider world.

For more information: http://eui.eu/rscas

Migration Policy Centre (MPC)

The Migration Policy Centre (MPC) is part of the Robert Schuman Centre for Advanced Studies at the European University Institute in Florence. It conducts advanced research on the transnational governance of international migration, asylum and mobility. It provides new ideas, rigorous evidence, and critical thinking to inform major European and global policy debates.

The MPC aims to bridge academic research, public debates, and policy-making. It proactively engages with users of migration research to foster policy dialogues between researches, policy-makers, migrants, and a wide range of civil society organisations in Europe and globally. The MPC seeks to contribute to major debates about migration policy and governance while building links with other key global challenges and changes.

The MPC working paper series, published since April 2013, aims at disseminating high-quality research pertaining to migration and related issues. All EUI members are welcome to submit their work to the series. For further queries, please contact the Migration Policy Centre Secretariat at migration@eui.eu

More information can be found on: http://www.migrationpolicycentre.eu/

Abstract

We devise an integrated estimate of country-to-country cross-border human mobility on the basis of global statistics on tourism and air passenger traffic. The joint use of these two sources allows us to (a) test for their relative contribution, and (b) correct for their limitations to the estimate of global mobility by combining them. The two sources are adjusted and merged following simple procedures. The resulting dataset, which covers more than 15 billion estimated trips over the years 2011 to 2016, promises to be a systematic and comprehensive resource on transnational human mobility worldwide. In this paper, we illustrate the data characteristics and transformations adopted in creating this dataset. First applications are explored, and its remaining limits are discussed.

Keywords

Transnational human mobility, travel, globalization, network data, tourism, air traffic.

1. Introduction

The increase in the cross-border mobility of persons—with differing reasons and objectives—is a hallmark of the current age of human history. In the face of this spectacular social trend, which is in place since at least the 1950s, there is a surprising dearth of systematic information detailing the size of travel flows across countries. The Global Mobilities Project (GMP) at the European University Institute's Migration Policy Centre (MPC) intends to fill this gap by addressing different dimensions of transnational mobilities (Recchi 2017).[1] As regards the volume and directions of mobility flows, we capitalize on two of the most comprehensive sources of transnational human movements at the global scale:

1. Data on tourism, i.e., cross-border visits that include an overnight stay (*nota bene*: not necessarily for leisure), from the World Tourism Organization (UNWTO);
2. Data on cross-border air passenger traffic from Sabre, a travel industry company.

Being conceived and collected for different purposes, both sources, taken individually, have clear limitations when used in the attempt to provide insights into human global mobility. The data on tourism is incomplete in that people moving between countries for reasons other than tourism (in particular, returning residents) are not included. It is also sometimes distorted because visitors from countries with few departures are not counted since their travel origin does not show up in the receiver country's tourism statistics. The data on air passenger traffic, in turn, does not factor in people who do not travel by airplane. In particular, journeys between neighboring countries, where cross-border mobility is particularly high (Deutschmann 2016), are likely to be severely underestimated since people often use car, railway, or bus transportation rather than flights. We propose to remedy these systematic biases by combining the two data sources, producing more reliable estimates of cross-country human mobility globally.

In this paper, we first make general considerations about the composition of transnational mobility flows in the two sources and give an overview of the procedures followed to combine them (section 2). We describe these procedures in more detail in section 3. Section 4 highlights some findings derived from first explorations of the newly created dataset. In the conclusion (section 5), we outline some pending limitations and advocate the use of the novel dataset to study transnational human mobility empirically in social science research.

2. Understanding the composition of transnational mobility flows

Our aim is to have robust estimates of the absolute number of yearly travels from and to every country worldwide.[2] In formal terms, we set out to measure the volume of cross-border travels T across all pairs of sovereign states a, b, c, … n on the planet. Such travels are carried out by both non-residents (NR) and residents (R) of receiving countries and take place by *air* (flights) or by *land/water* transportation

[1] We use the term 'transnational' in the meaning it has in the field of international relations, where it is employed to describe any movement by non-state actors that spans across national borders (Nye and Keohane 1971). We are aware that in the field of migration studies 'transnational' has a more demanding meaning that involves the regular movement of the *same* individuals across certain borders (Wimmer and Glick Schiller 2002). Following the latter tradition, it would be equally justified to speak about *international* mobility.

[2] Conceptually, migrants and asylum-seekers are excluded from our estimates, even though we cannot rule out that some 'visitors' may overstay their travels and thus become migrants and asylum-seekers. More on the issue in the Conclusions (section 5).

(trains, buses, cars and private road vehicles, boats, ferries and ships),[3] that we indicate respectively with exponents ^{A} and ^{L}. Therefore:

$$T_{a\to b} = NR^A_{a\to b} + R^A_{a\to b} + NR^L_{a\to b} + R^L_{a\to b}$$

Unfortunately, no existing source contains information on all four factors simultaneously. The original tourist files include only $NR^A_{a\to b} + NR^L_{a\to b}$, i.e., they register tourist *arrivals* in destination countries, but not tourists returning to their countries of origin. Air traffic statistics include $NR^A_{a\to b} + R^A_{a\to b}$, i.e., air passengers only.[4] Thus, both datasets are suboptimal as they systematically exclude $R^L_{a\to b}$. We expect the two datasets to be strongly correlated, because they share the same core component: $NR^A_{a\to b}$. They should diverge only when $R^A_{a\to b}$ and/or $NR^L_{a\to b}$ are large and/or not correlated.

The original UNWTO tourist files, however, also record residents of b going from b to *a with all transportation means*, that is $R^A_{b\to a}$ and $R^L_{b\to a}$. If we imagine that these people return to their country of residence in the same year of their outbound travel, we can count them as part of $R^A_{a\to b}$ and $R^L_{a\to b}$. We can thus assume that $R^A_{a\to b} + R^L_{a\to b} = R^A_{b\to a} + R^L_{b\to a}$. This assumption falls short of a small proportion of travellers who: a) travel by the end of the year and come back in the following calendar year, or b) resettle abroad. As for a), given the overall constancy of travel flows, we can maintain that these travellers are offset by similar travellers twelve months earlier. As for b), these travellers are migrants. A comparison of migration flows (as estimated by Abel and Sander 2014) and global tourist flows (based on Deutschmann 2016) shows a 1/150 relationship. That is, migrant travels correspond to about 0.6 percent of tourist travels, which is therefore the approximate overall size of error we introduce in our tourism estimates with this assumption.[5] We therefore revise the original UNWTO tourism data to build a yearly matrix of tourists/visitors travelling from a to b which also includes (returning) travellers from b who moved to a:

$$T^{revised}_{a\to b} = NR^A_{a\to b} + NR^L_{a\to b} + R^A_{b\to a} + R^L_{b\to a}$$

Hereafter, we will call this the GMP-revised tourism data [1]. Its creation is described in detail in section 3.1.

As for the air passenger data, which we use in its KCMD-revised form [2] (see explanation below), we assume that they tend to be lower than the revised tourism data [1] because travellers also move with other transportation means. However, [1] and [2] should converge the larger the distance between origin and destination as air travel tends to become the exclusive means of transportation at long distances. This distance-mediated relationship between [1] and [2] leads us to transform the air passenger data. We compute an estimate of transnational mobility [3] that adjusts [2] by a factor that accounts for the distance between countries. The formal procedure to estimate [3] is described in section 3.3.

In a final step, we combine the two revised sources, [1] and [3], to create an integrated dataset on global transnational mobility. As we hold that both [1] and [3] tend to underestimate actual mobility flows, our final estimate is always the largest of the two when we have both information—that is, either [1] or [3]. When we lack [3], we take [1], and vice versa.

---

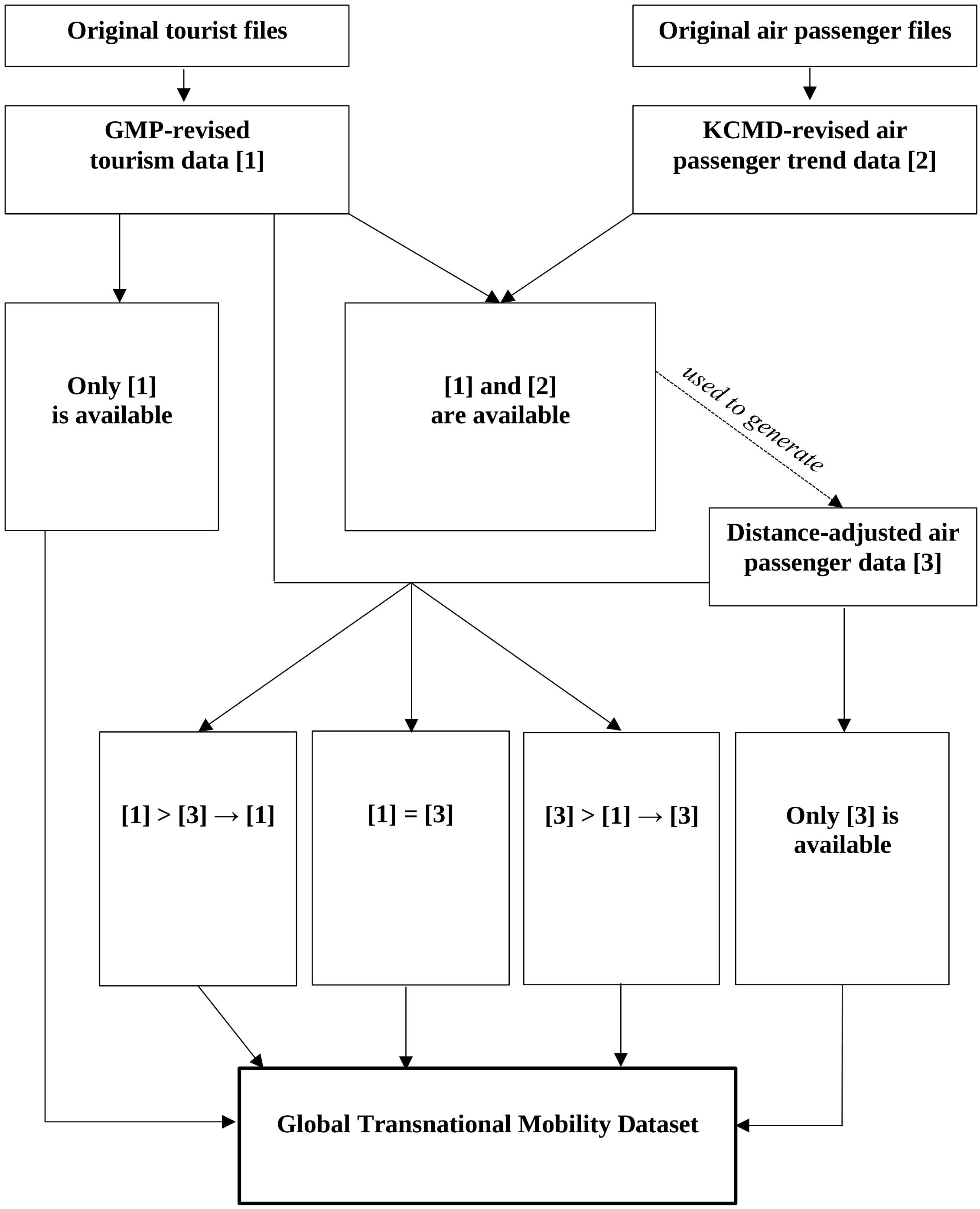
3 Other statistically marginal forms of mobility (by foot or bike, for instance) are also included, provided they take place legally (i.e., they are registered). Unregistered or illegal border crossings are in fact left out by default from tourism and air traffic statistics, and, as a consequence, from our estimates.

4 Note that air traffic statistics do not allow us to distinguish between these two components since they are based on the location of the airport of origin and destination, not on the residence or nationality of the traveller.

5 As our final estimate of global transnational mobility provides higher figures, the actual migrant/travel ratio is likely to be even lower, with migrant flows corresponding to about 0.001 percent of travel flows (see section 4, Figure 6).

Figure 1 provides an overview of this procedure. The individual steps are described in more detail in the following sections. The resulting final dataset covers 196 sender and receiver countries, generating a symmetric matrix of 38,220 cases (i.e., country pairs) per year. For the entire 2011-2016 period, about 9.5 billion trips (approx. 61 percent) are ultimately derived from [1] and 6 billion trips (approx. 38 percent) from [3]. Overall, 12.0 percent of cells are empty, which can mean either a total absence of transnational mobility between these countries (most likely in the case of pairs of small and distant nations) or missing data. The Global Transnational Mobility Dataset covers an estimated total of 15.7 billion trips.

Figure 1. Overview of the data composition



3. Constructing the dataset

In the following subsections, we outline in more detail how we handled the raw data and proceeded until the creation of the final Global Transnational Mobility Dataset. We first describe the creation of the GMP-revised tourism data (section 3.1). Second, we bring the KCMD-revised air passenger trend data in (section 3.2). Third, we introduce the correction factor that adjusts the latter source, taking geographic distance into account (section 3.3). Finally, we describe the merging and finalization of the dataset (section 3.4).

3.1 Creating the GMP-revised tourism data [1]

Our first and primary source, the UNWTO tourism data, was obtained by the Global Mobilities Project (GMP) of the EUI's Migration Policy Centre (MPC) from the UNWTO as a set of files containing yearly flows from 1995 to 2016 for all sovereign countries and dependent territories worldwide (UNWTO 2015).[6] The original data contains 219 excel files, one per receiver country/territory. To create a unified, standardized, and usable dataset (hereafter the GMP-revised tourism data), we took the following steps:

Step 1: Prioritizing the different UNWTO operationalizations of 'arrivals'

The country-to-country flow data on arrivals is reported in eight different categories in the UNWTO data (Table 1). The UNWTO defines arrivals—and describes its sources—as follows:

> *Arrivals* data measure the flows of international visitors to the country of reference: each arrival corresponds to one inbound tourism trip. If a person visits several countries during the course of a single trip, his/her arrival in each country is recorded separately. In an accounting period, arrivals are not necessarily equal to the number of persons travelling (when a person visits the same country several times a year, each trip by the same person is counted as a separate arrival).
>
> Arrivals data should correspond to *inbound visitors* by including both tourists and same-day non-resident visitors. All other types of travelers (such as border, seasonal and other short-term workers, long-term students and others) should be excluded, as they do not qualify as visitors. Data are obtained from different sources: administrative records (immigration, traffic counts, and other possible types of controls), border surveys or a mix of them. If data are obtained from accommodation surveys, the number of guests is used as estimate of arrival figures; consequently, in this case, breakdowns by regions, main purpose of the trip, modes of transport used or forms of organization of the trip are based on complementary visitor surveys. (UNWTO 2015, p. 9)

To include as many cases as possible in the unified dataset, we use all eight 'arrivals' categories shown in Table 1, in order of preference. This preference order is justified on the basis of a number of assumptions and compromises that are discussed in the Appendix.

Step 2: Creating a unified excel file

We then created a unified excel file that contains the relevant country-to-country flow data to all countries for which this information was available.[7] In doing so, we exclude several 'odd' sender categories, such as 'Other countries of the world', which cannot readily be included in a country-to-country flow matrix. Details about this procedure and its consequences are described in the Appendix.

[6] At UNWTO, we thank Jacinta Mora for facilitating our access to these tourism statistics.

[7] There are 18 countries that are part of the UNWTO data collection that do not report country-to-country flow data. This means they may be part of the full tourism dataset as senders of tourists but not as receivers. They are: Afghanistan, Bonaire, Djibouti, Equatorial Guinea, Eritrea, Gabon, Ghana, Guinea-Bissau, Liberia, Libya, Mauritania, Saba, Sao Tome and Principe, Sint Eustatius, South Sudan, Syrian Arabic Republic, Turkmenistan, and United Arab Emirates.

Table 1. Categories in the UNWTO dataset

Code	Description	Preference
112	Arrivals of non-resident tourists at national borders, by country of residence	1st
111	Arrivals of non-resident tourists at national borders, by nationality	2nd
122	Arrivals of non-resident visitors at national borders, by country of residence	3rd
121	Arrivals of non-resident visitors at national borders, by nationality	4th
1912	Arrivals of non-resident tourists in all types of accommodation establishments, by country of residence	5th
1911	Arrivals of non-resident tourists in all types of accommodation establishments, by nationality	6th
712	Arrivals of non-resident tourists in hotels and similar establishments, by country of residence	7th
711	Arrivals of non-resident tourists in hotels and similar establishments, by nationality	8th

Step 3: Adding returning residents

In line with the considerations made in section 2, we add the returning residents $R^{A}_{b\rightarrow a} + R^{L}_{b\rightarrow a}$, to the incoming non-residents $NR^{A}_{a\rightarrow b} + NR^{L}_{a\rightarrow b}$ to obtain a more complete picture of human mobility across borders. In doing so, we effectively double the number of trips in the tourism dataset. Furthermore, the matrix becomes symmetric, i.e., mobility flows are now, by necessity, the same in both directions ($T^{revised}_{a\rightarrow b} = T^{revised}_{b\rightarrow a}$). After this step, we have obtained the GMP-revised tourism data [1].

3.2 Bringing the KCMD-revised air passenger trend data [2] in

The second source is the dataset on global air passenger traffic in the 2011–2016 period collected by a private travel industry company, Sabre (2014). The dataset contains information on the number of air passengers per month, traveling between airports. Here, we draw on a simplified and reduced version created by researchers at the European Commission's Knowledge Centre for Migration and Democracy (KCMD) that represents the yearly trend between countries (henceforth KCMD-revised air passenger trend data [2]). This version was generated through a time-series decomposition that dissects the raw overall air passenger flow between two countries into a trend component, a seasonal component, and a residual component (Gabrielli et al. 2019). In the KCMD-revised air passenger trend data [2] used here, the monthly trend data is aggregated to yearly averages. The data is available for the years 2011 to 2016.

We merge the two datasets [1] and [2] using ISO 3166-1 alpha-3 country codes. In line with the considerations made in section 2, we hypothesize:

a) [1] to be on average larger than [2], as it includes both air passengers and land/water travellers;
b) [1] and [2] to be highly correlated, since many travellers use flights to cross borders;
c) [1] and [2] to be more strongly correlated as the distance between country pairs increases, since people are more likely to use air transportation at longer distances.

All three hypotheses hold empirically. As expected, tourism figures based on [1], reporting cross-border trips with all transportation means, tend to be higher than air passenger figures based on [2], reporting journeys that take place with flight transportation only. The exceptions are mainly countries receiving by plane a number of returning residents or nationals exceeding the number of non-national visitors (de facto, out-migration countries with little incoming tourism). Table 2 shows the distribution of the deviations between the two data sources across cases (i.e., country pairs), by year. Negative values denote that there are more tourists than air passengers; positive values denote that there are more air passengers than tourists travelling between a pair of countries. The average median (50th percentile) across years is -2,410 trips, and even at the 75th percentile of cases, there are still more tourists than air passengers (-85 trips). Table 2 also reveals that, as the distribution is quite stable over time, the divergence between the two sources is no coincidence, but does indeed reflect the structural difference described above in hypothesis (a).

Table 2. The distribution of the difference between tourists and air passengers

Percentiles	2011	2012	2013	2014	2015	2016
Min	-89,300,000	-89,800,000	-89,400,000	-90,200,000	-92,400,000	-93,400,000
1%	-3,918,997	-4,064,395	-4,002,791	-4,361,469	-3,865,980	-4,136,718
5%	-514,371	-581,089	-661,828	-655,484	-569,920	-643,928
10%	-192,821	-212,287	-235,487	-232,265	-183,901	-218,354
25%	-22,009	-27,635	-30,651	-28,778	-24,436	-28,451
50%	-1,997	-2,536	-2,924	-2,493	-2,189	-2,323
75%	-63	-113	-126	-94	-56	-55
90%	1,770	1,220	998	1,371	1,480	4,097
95%	11,824	10,775	8,400	10,081	10,992	28,604
99%	131,253	140,405	109,720	113,494	140,005	257,340
Max	1,137,767	834,788	1,070,940	1,191,830	1,396,962	2,525,211
Obs.	5,359	5,771	5,649	5,653	5,779	5,262
Mean	-210,505	-219,209	-232,735	-232,250	-224,670	-243,573
Std. Dev.	2,175,910	2,132,248	2,178,926	2,221,919	2,251,131	2,393,686
Skewness	-30	-30	-28	-28	-29	-27
Kurtosis	1,105	1,131	1,048	1,020	1,043	939

Note: Negative values denote that there are more tourists than air passengers; positive values denote that there are more air passengers than tourists travelling between a pair of countries.

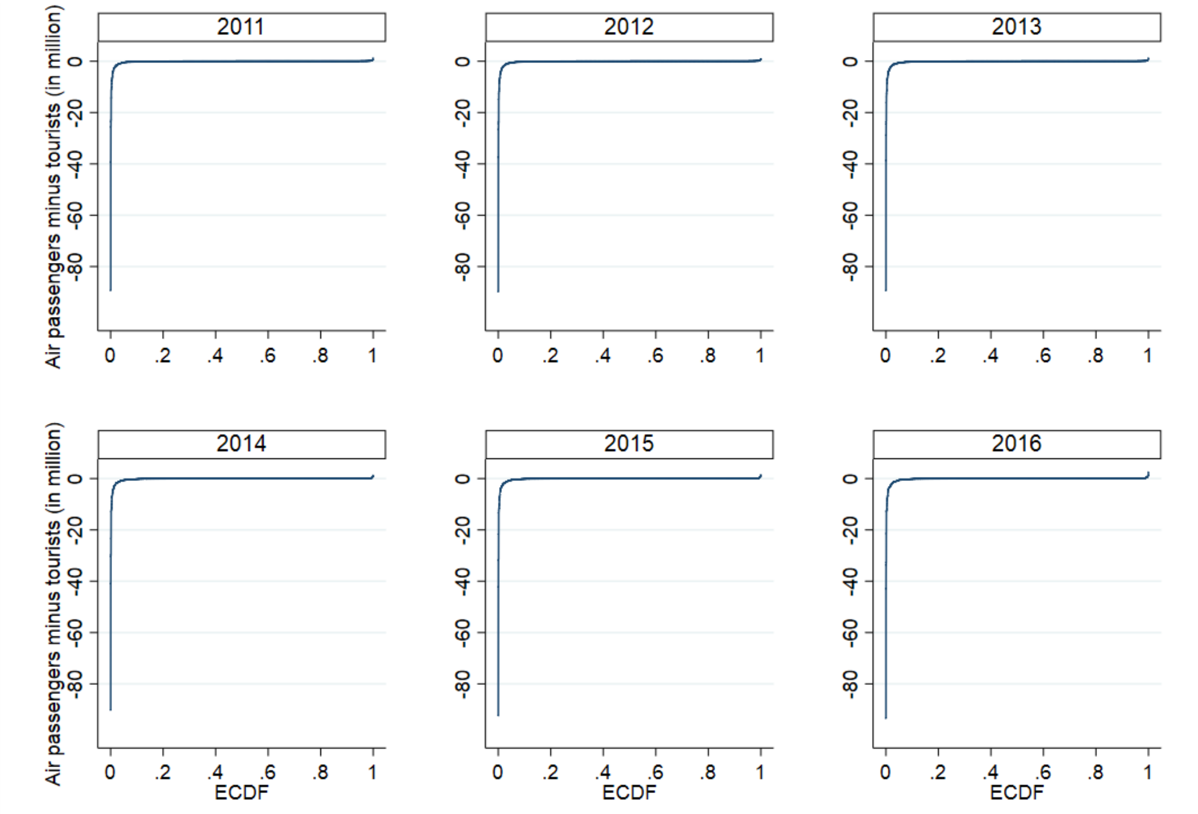
Figure 3 shows the relation between the tourist-air passenger discrepancy and geographic distance (based on CEPII's GeoDist dataset [Mayer and Zignago 2006]). A clear pattern emerges: there are sizeable discrepancies at short geographic distances only. The most extreme negative deviations (i.e., a lot more tourists than air passengers) are Hong Kong ↔ China (89-93 million, depending on year and direction), Macao ↔ China (37-43 million), United States ↔ Mexico (30-34 million), and Germany ↔ Poland (26-33 million). As Figure 3 clearly shows, extreme cases consistently cluster together over time (the rings of different colors represent the different years). This suggests that these discrepancies are not random but systematic and meaningful. The inspection of specific cases with the highest negative[8] deviations helps understand the rationales of the discrepancies, which can overlap and reinforce each other:

1. Mobility between *nearby countries*: tourists exceed air passengers because many people move across borders with land (train, car, bus) or water (ferry, ship) transportation. Examples include the four extreme outlier country pairs tagged in Figure 3.
2. *Grand tour tourism*: Here, people fly to one country (e.g., from the U.S. to the Netherlands), and then go by car or train to other countries (e.g., France). In France, they are counted as tourists (e.g., through hotel registration data) but not as air passengers.

[8] In fact, there are few exceptional cases in which air passengers are in larger numbers than registered tourists. These are mostly distant countries with large contingents of migrants or returning nationals (who are not registered by tourism statistics) but relatively modest inflows of other visitors (e.g., India and Oman).

While rationale (2) is difficult to deal with (see the remaining limitations described in section 5), we treat rationale (1) by creating a correction factor that takes distance into account.

Figure 2. Cumulative distributions of the difference between the GMP-revised tourism dataset



[1] and the KCMD-revised air passenger trend dataset [2]

Note: ECDF = Empirical cumulative distribution function

3.3 Creating the distance-adjusted air passenger data [3]

The goal here is to adjust the KCMD-revised air passenger trend data [2] to correct for the fact that it underestimates mobility at short distances due to the use of alternative transportation means. To do so, we draw on the distance (in km) between country pairs. Our correction factor is specified as:

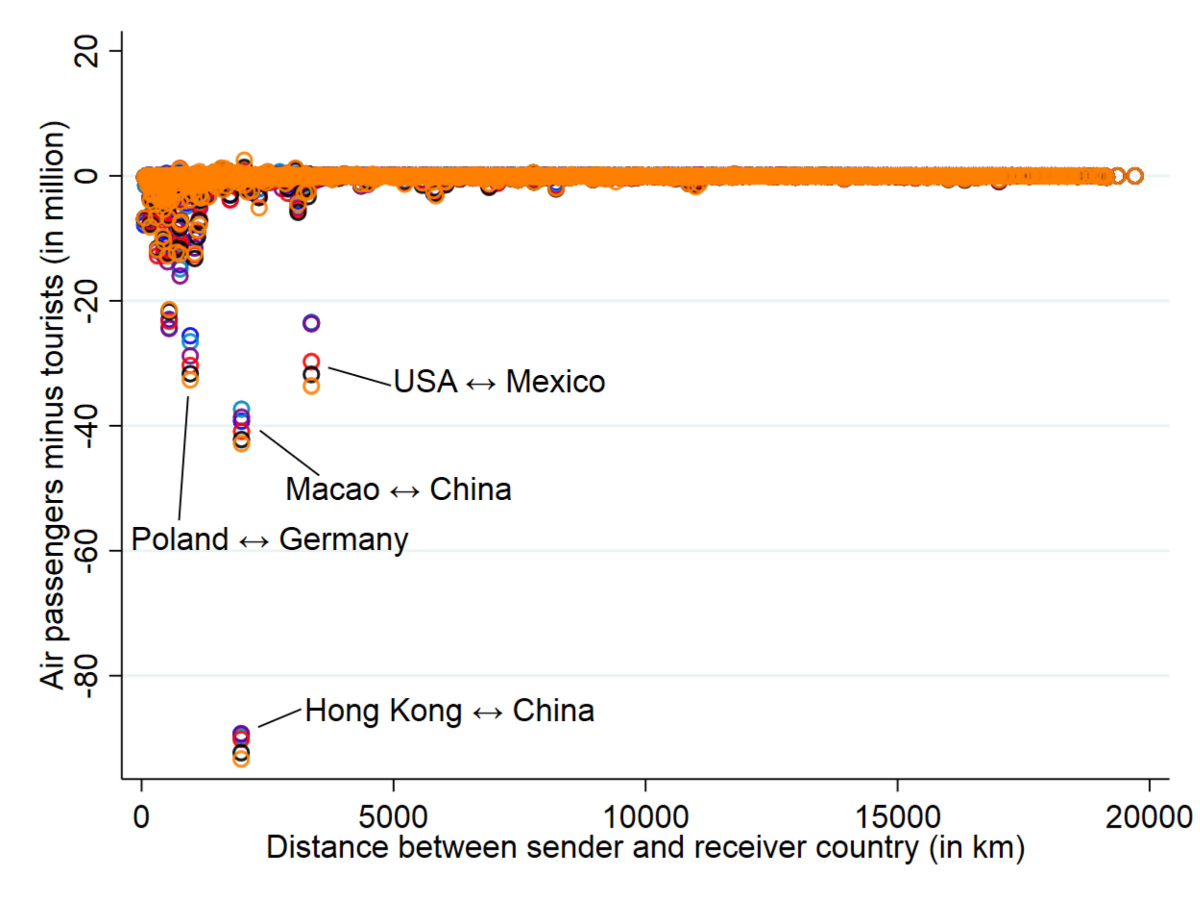
$$\left(\frac{k_{max}}{k_{A\leftrightarrow B}}\right)^{1/c}$$

where k_{max} is the maximum possible distance between two countries, in this case 19,951.16 km (the distance between Paraguay and Taiwan), and $k_{A\leftrightarrow B}$ is the empirical distance between two countries *A* and *B*, based on CEPII's GeoDist dataset (Mayer and Zignago 2006). The parameter *c* is chosen so that it maximizes the correlation *r* between the GMP-revised tourism data [1] and the KCMD-revised air passenger trend data [2].[9] The rationale behind this is the assumption that [1] is not biased in terms of

[9] We combine data from all available years and exclude cases with more than 10 million trips to reduce the influence of these outliers on the calculations. On average, 31 cases are ignored per year (0.08 percent of the total).

distance. Distance-adjusting [2] so that its correlation with [1] is maximized should thus lead to the best possible correction factor.

Figure 3. The relation between geographic distance and divergences between the GMP-revised tourism dataset [1] and the KCMD-revised air passenger trend dataset [2]



Note: Different colors denote different years. Distance is obtained from Mayer and Zignago (2006)

Figure 4. Adjusting the distance-based correction factor for the KCMD-revised air passenger trend data to maximize the fit with the GMP-revised tourism data

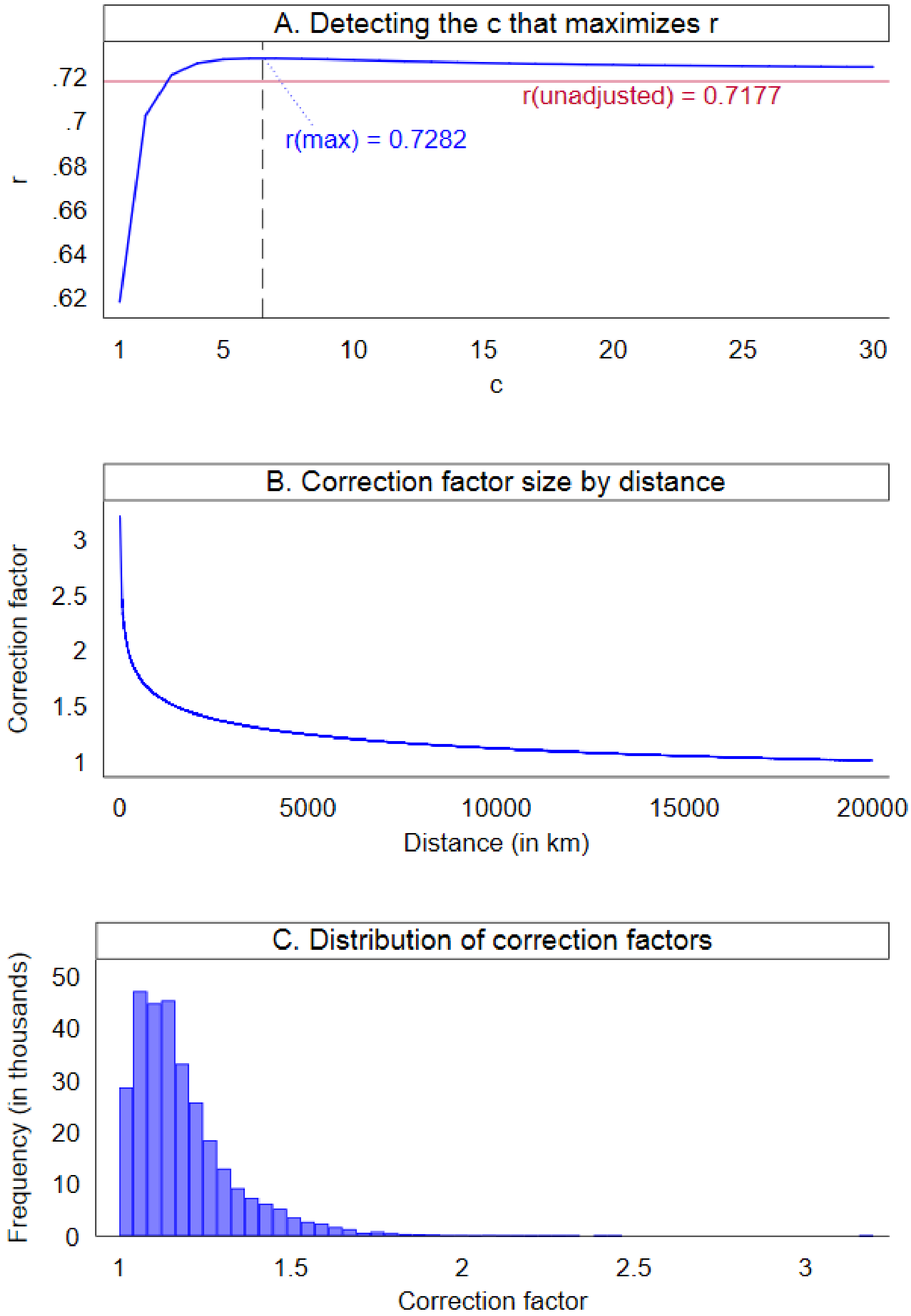
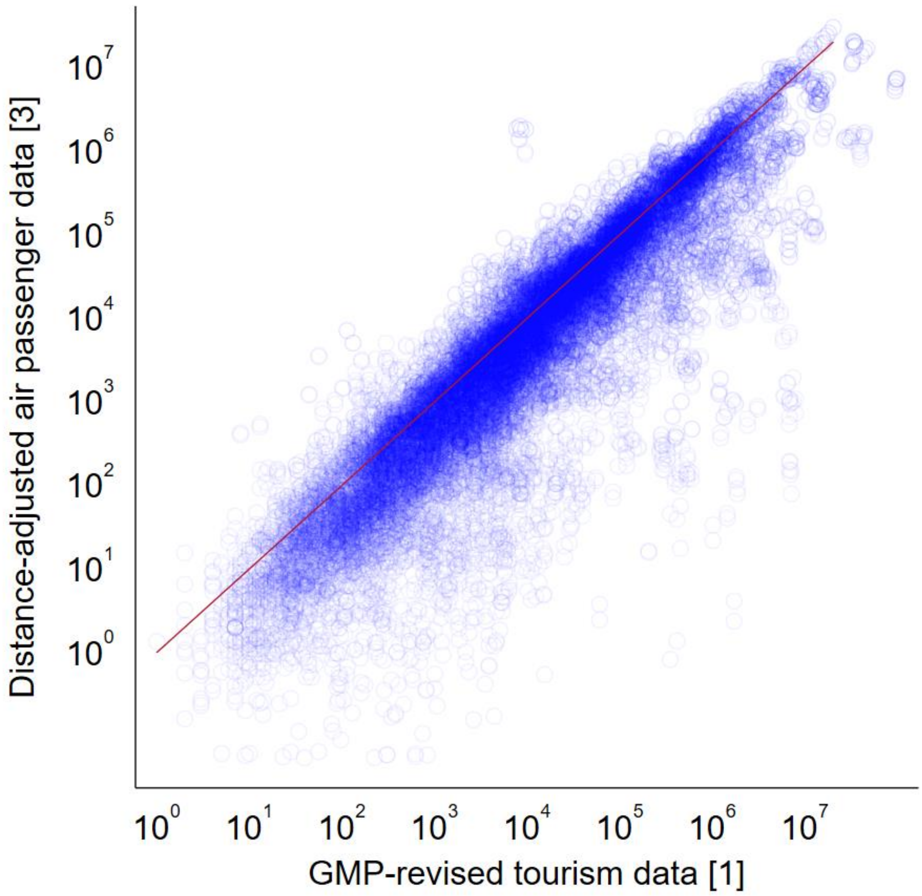


Figure 5. The correlation between the distance-adjusted air passenger data [3] and the GMP-revised tourism data [1]



3.4 Creating the Global Transnational Mobility Dataset

In the final step, we merge the two revised data sources. As we hold that both the GMP-revised tourism data [1] and the distance-adjusted air passenger data [3] individually tend to under-estimate actual mobility flows (see section 2), our final estimate is always the largest of the two when we have both information—that is, either [1] or [3]. When we lack [3], we take [1]; and vice versa. As final steps, we:

- Round decimals (non-integer estimates can occur due to the time-series decomposition applied by Gabrielli et al [2019] and the correction factor introduced in section 3.3).
- Add missing full country names and information on the world region a country is situated in based on the United Nations classification (drawing on Duncalfe [2018]).
- We exclude countries for which, after the merging procedure, no information was available.[10] Consequently, the dataset is reduced to the set of 196 countries used when creating the unified UNWTO dataset.

[10] Countries and territories excluded are: Aruba, Anguilla, Cocos Islands, Cook Islands, Christmas Islands, Western Sahara, Falkland Islands, Faroe Islands, Guadeloupe, Grenada, Greenland, French Guiana, Montenegro, Northern Mariana Islands, Montserrat, Martinique, New Caledonia, Norfolk Islands, Pitcairn, Puerto Rico, French Polynesia, Reunion, Saint Helena, Saint Pierre and Michelon, Serbia, Tokelau, Taiwan, Wallis and Futuna Islands.

The resulting Global Transnational Mobility Dataset can be explored on an interactive world map at the KCMD Dynamic Data Hub (https://bluehub.jrc.ec.europa.eu/migration/app/index.html; browse ‘Datasets’ – ‘Mobility (JRC)’ – ‘Estimated Trips (KCMD-EUI)’). More information on the website of the Migration Policy Centre of the EUI (http://www.migrationpolicycentre.eu/globalmobilities/). The dataset can be requested for scientific research by email (GMPdataset@eui.eu). It contains the following variables:

Table 3. Variables contained in the Global Transnational Mobility Dataset

Name	Description
source_name	Name of the country of origin
target_name	Name of the country of destination
source_iso3	ISO 3166-1 alpha-3 code of the country of origin
target_iso3	ISO 3166-1 alpha-3 code of the country of destination
year	Year, ranges from 2011 to 2016
estimated_trips	Estimated trips
dist	Geographic distance
source_region	Region of the country of origin
target_region	Region of the country of destination
source_subregion	Sub-region of the country of origin
target_subregion	Sub-region of the country of destination

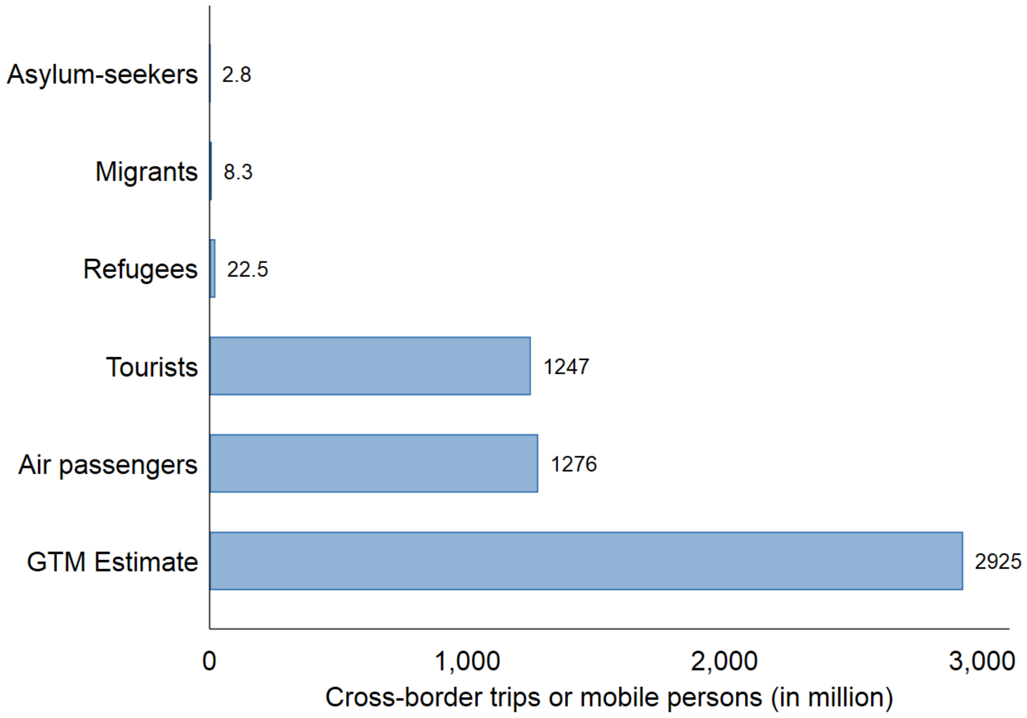
Note: Geographic distance is obtained from CEPII’s GeoDist dataset (Mayer and Zignago 2006). Regions and subregions are based on the UN M.49 GeoScheme.

4. Exploring the dataset: some first insights

The Global Transnational Mobility Dataset covers 196 sender and receiver countries and, through the integration of two different sources, is more comprehensive than all pre-existing information on worldwide cross-border mobility. This is illustrated in Figure 6, which displays the estimates of mobility given by several sources. According to UNHCR, there were 2.8 million new asylum-seekers crossing borders globally in 2016. The number of yearly migrant flows is very difficult to establish, but according to one estimate, it could be around 8 million people per year.[11] The global stock of refugees is estimated to be 22.5 million for 2016 (UNHCR 2016). In the original UNWTO tourism files, around 1.3 billion tourist trips are recorded. A similar number is obtainable from the KCMD-revised air passenger trend data. According to our new dataset, there were about 2.9 billion cross-border trips in 2016.

[11] This figure is based on Abel and Sander (2014) and is obtained by dividing the estimate of global migration flows from the mid-2005 to mid-2010 period by 5. Estimates for more recent years are unfortunately unavailable.

Figure 6. Comparison between estimates



Note: Sources: Tourists: UNWTO (2016); migrants: Abel and Sander (2014) (estimate of global migration flows from the mid-2005 to mid-2010 period divided by 5); refugees and asylum-seekers: UNHCR (2015); air passengers: KCMD-revised air passenger trend data. Note that the unit differs between sources: asylum-seekers, migrants and refugees are mobile *persons*, whereas tourists, air passengers and Global Transnational Mobility (GTM) data are recorded in cross-border *trips*.

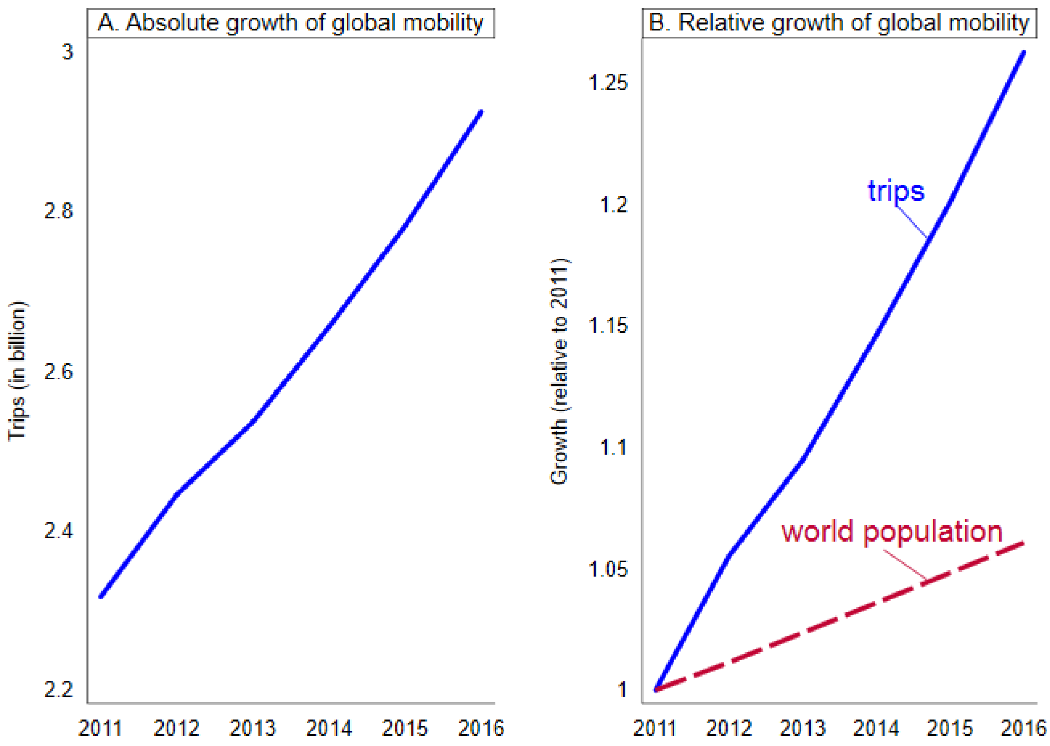
While we leave to future research the full exploitation of its potential, also in conjunction with other datasets, a preliminary exploration of the Global Transnational Mobility Dataset offers several major insights that are detailed in this section.

4.1 Worldwide transnational mobility is rapidly increasing over time

Figure 7 shows that during the time frame under study, 2011 to 2016, transnational human mobility increased dramatically. In absolute terms, the number of estimated trips increased from about 2.3 billion in 2011 to about 2.9 billion in 2016 (Figure 7A). As Figure 7B reveals, this growth is much larger than the growth in world population, indicating that collectively, humanity has indeed become more transnationally mobile. In this regard, transnational mobility is developing similarly as cross-border communication, but differently from migration, which has not grown significantly faster than the world population (Deutschmann 2016).

This development raises questions for many fields of inquiry, like the environmental consequences, the potential spread of epidemics, the emergence of global systemic risks (Centeno et al. 2015) and, from a sociological perspective, the social inequalities in access to these increased mobility chances. The latter issue is briefly touched upon in the following section.

Figure 7. Absolute and relative growth of global mobility



Note: The graphs are based on the Global Transnational Mobility Dataset (trips) and World Bank (2018) population data.

4.2 Transnational mobility tends to cluster within world regions

Figure 8A shows the mobility (in million trips) within world regions, using the United Nations M.49 Geoscheme as a base for assigning countries to regions. We find that Europe is the region with the highest number of intraregional trips, followed by Asia. The Americas are behind, and the smallest number of trips occur within Africa and Oceania. Over time, between 2011 and 2016, intraregional mobility grows strongly in Europe and Asia. The Americas see a smaller increase and mobility in Africa and Oceania looks much more stable in comparison. There is thus no clear catch-up effect. Rather the divergence between regions in terms of intraregional mobility seems to widen over time.[12]

Interregional mobility can be studied by either taking the outgoing mobility *from* a specific region (Figure 8B) or the incoming mobility *to* a specific region (Figure 8C) into account. Both strategies yield very similar outcomes. In both cases, interregional mobility is far less common than intraregional mobility, at least for Europe, Asia, and the Americas (cf. Figure 9 and its discussion below). Also note that the order between regions is the same in terms of intra- and interregional mobility.

Figure 9 allows us to take a closer look at the ratio of intra- to interregional mobility by region. This could be described as a measure of relative regionalism (Deutschmann 2017). This indicator reveals a

[12] Note that this simple measure may not be the best one to study how regionalized mobility actually is. It is well possible that within Europe, for example, the high number of trips is driven by a subset of country pairs and that others participate very little in the intraregional network of transnational human mobility. Deutschmann (2017) proposes to use density-based measures as an alternative that allows to take into account between how many country pairs in a region meaningful amounts of mobility exist. Moreover, more sophisticated analyses would have to consider the varying population sizes of regions.

very similar picture regardless of whether incoming or outgoing mobility is used as a measure. In both cases, intraregional mobility is more than 5 times more likely to occur than interregional mobility in the case of Europe, more than 4 times in the case of Asia, and almost 3 times in the case of the Americas. In the case of Africa intraregional mobility is basically as likely as interregional mobility, and in Oceania, intraregional mobility is even half as likely as interregional mobility.

Note, however, that this comparison may be seen as ‘unfair’ since the pool of potential connections is obviously much larger in the case of interregional mobility than in the case of intraregional mobility. A more sophisticated and ‘just’ comparison (which goes beyond the scope of this paper) would be to compare intraregional mobility to mobility towards specific other world regions. Past research has found that when this is done, mobility also tends to cluster within Africa and Oceania (Ibid.).

In any case, Figures 8 and 9 highlight the extreme stratification of the chance to engage in transnational mobility at the global scale. Transnational mobility within Europe is about 20 times the amount of mobility within Africa, in spite of the much larger population of the latter continent. This global inequality in mobility chances has important sociological implications. For example, it has been shown that transnational human capital is an important resource that increases life chances (Gerhards et al. 2017). Furthermore, transnational mobility shapes world views, attachment to other countries and cosmopolitan attitudes (Mau et al. 2008; Helbling and Teney 2015; Kuhn 2015; Recchi 2015; Deutschmann et al. 2018; Recchi et al. 2019). While these consequences of unequal access to transnational mobility chances have mainly been studied from a European viewpoint so far, a global perspective is largely missing. The Global Transnational Mobility Dataset may prove a good starting point for future analyses in this direction. The next section digs a little deeper into this global stratification by looking at the relation between transnational human mobility and levels of prosperity.

4.3 Transnational mobility differs by levels of prosperity and country size

Figure 10 illustrates how transnational mobility differs by levels of prosperity and country size. Figure 10A shows a clear relation between a country’s outgoing trips and the national level of prosperity, measured as GDP per capita in purchasing power parity (World Bank data). The relation is relatively strong and significant, with a correlation coefficient of r = .63. Figure 10B shows a similar pattern for the relation between mobility and population size. Again, the correlation is quite high with r = .58. The three-dimensional graph in Figure 10C illustrates the relation between the three factors in combination. The distribution of dots, representing countries, follows a clear pattern, ranging from low GDP, small population and low mobility (blue dots at the bottom front corner) to high GDP, large population and high mobility (red dots in the upper back corner). These insights are not entirely new (e.g., Deutschmann 2016 and 2017), but are showcased in a clear and robust way by this novel dataset. Future research may engage in more complex analyses, taking a larger set of factors into account and building more comprehensive multivariate models to study the antecedents and consequences of transnational human activity worldwide.

Figure 8. Mobility within and between world regions

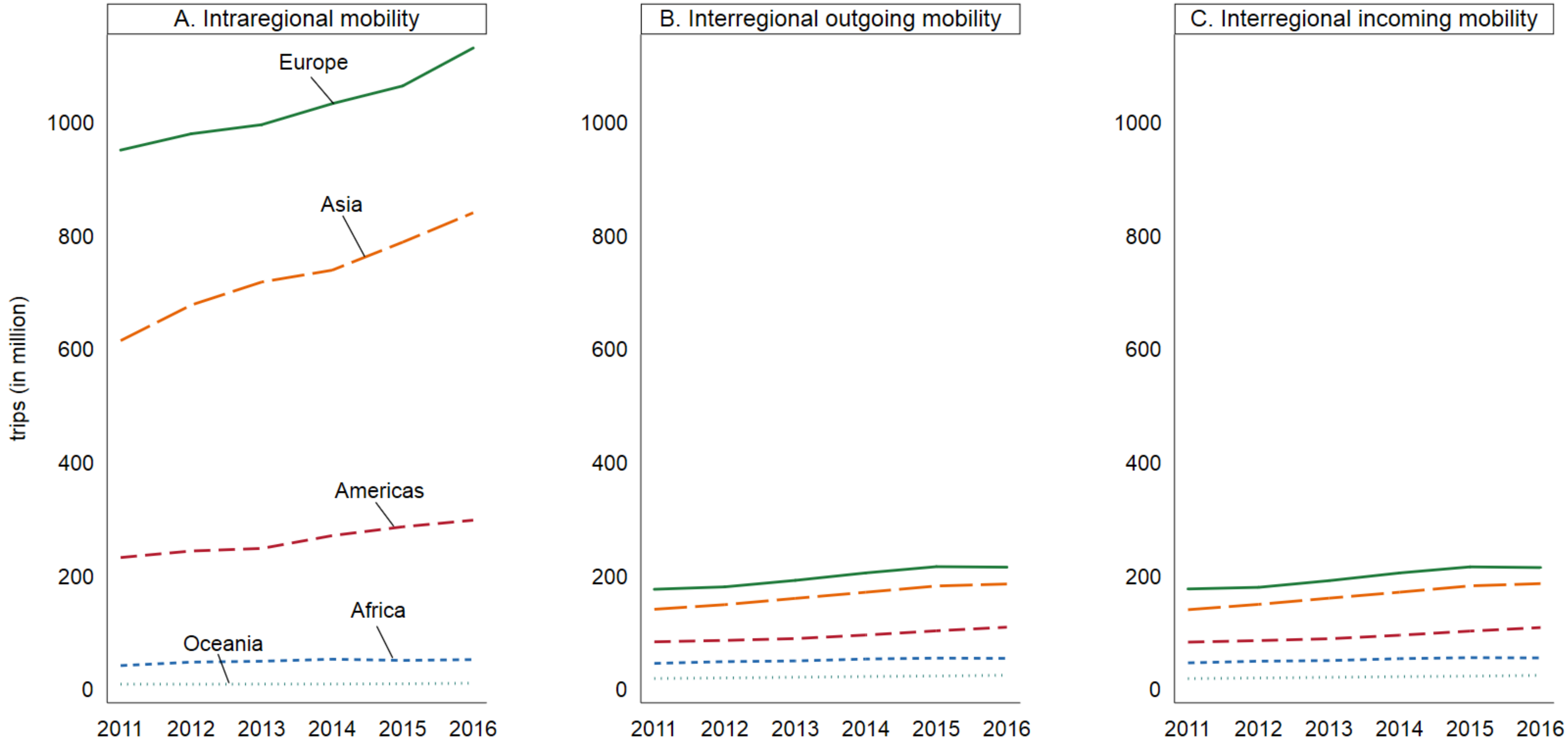


Figure 9. Relative regionalism, by region

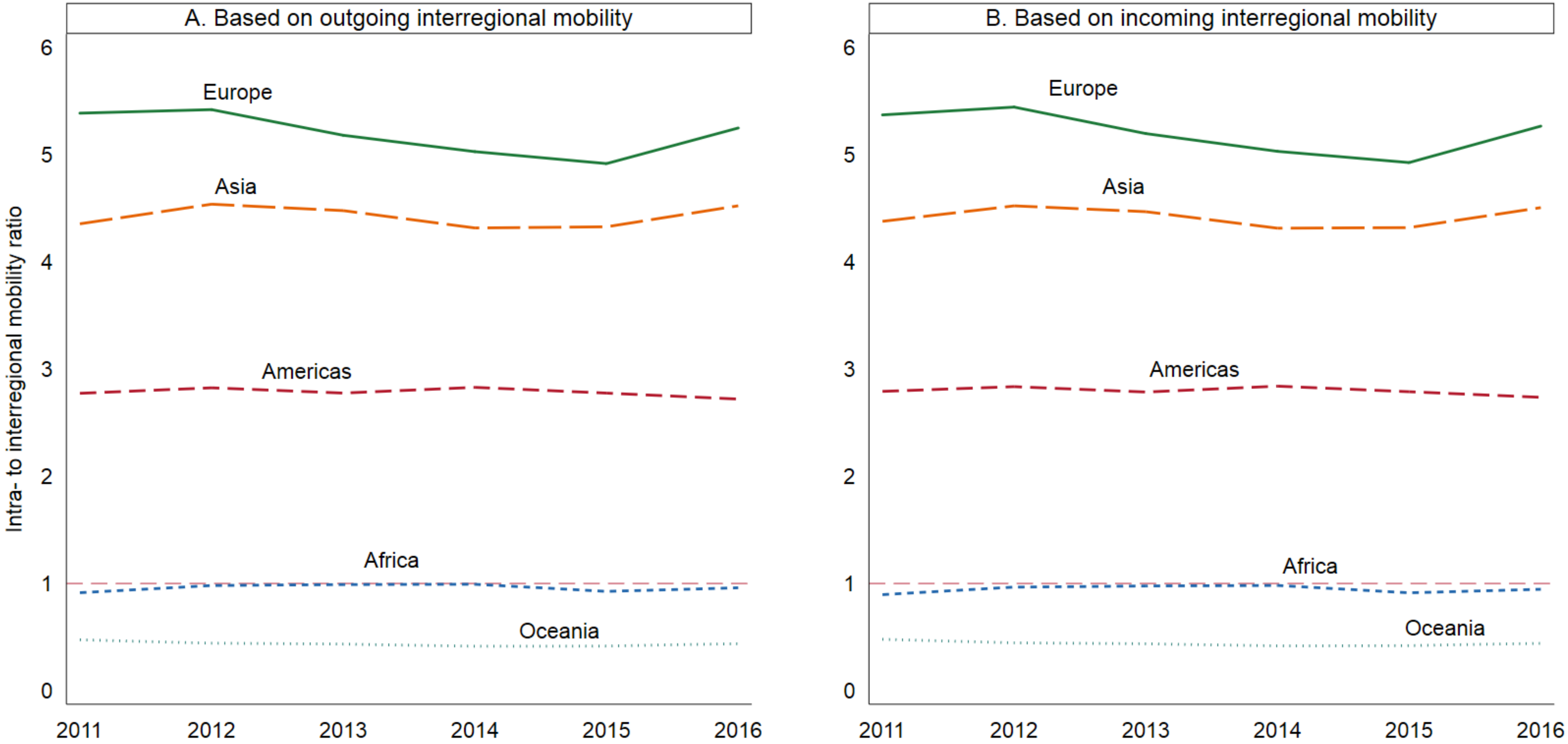
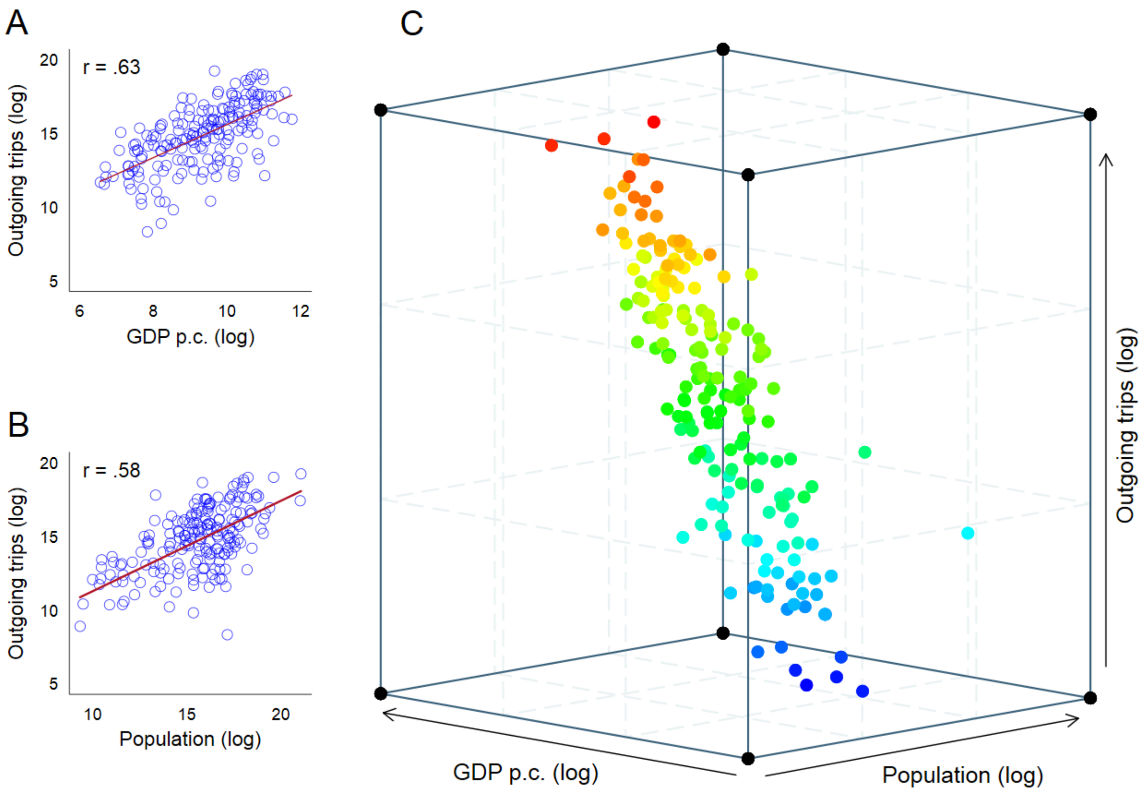


Figure 10. The relation between mobility, population size, and GDP per capita.



5. Conclusion

A spate of migration and asylum-seeking crises has been hitting the world since the turn of the 21st century. The globe is on the move but, in spite of their over-exposure in the media and public opinion, migrants and refugees constitute only a tiny portion of the whole number of people crossing borders daily. According to Abel and Sanders' (2014) estimates, there were less than 10 *million* worldwide migration episodes per year in the early 2010s worldwide. According to our estimate, yearly border-crossings come close to 3 *billion* globally. By providing estimates of the amount of such transnational mobility beyond migration, the Global Transnational Mobility Dataset—created as an outcome of this paper—facilitates the study of the volume, directions and change of country-to-country human mobility on a worldwide scale.

This paper has described the procedures by which we have reached these estimates. While we acknowledge that there is no single existing data source providing exact information on the number of people officially crossing national borders worldwide, we do find that the two more complete and reliable sources (data on tourism and data on air passengers) do show significant consistency and can be merged according to a few and relatively simple combination rules.

Focusing on yearly country-to-country flows of human *mobility* (whatever their duration), our dataset complements estimates of worldwide *migration* flows (Abel and Sanders 2014), which refer to stays abroad longer than 12 months, based on the conventional UN definition of migration. This dataset also advances previous usages of the UNWTO data (Reyes 2013; Deutschmann 2016 and 2017), capitalizing on an additional source and estimation methods. Finally, the Global Transnational Mobility Dataset parallels recent and alternative attempts at measuring population mobility with digital sources (State et al. 2013; Hawelka et al. 2014; Messias et al. 2016; Rango and Vespe 2017; Zagheni et al. 2017; Fiorio et al. 2017; Spyratos et al. 2018). Data triangulation across these digital estimates and ours may prove useful to test the comparability of outcomes obtained through such different approaches.

Several important limitations remain. The first issue concerns the existence of grand-tour tourism and open-jaw flights (see section 3.2). For instance, consider a traveler who goes on a round trip to Southeast Asia from Italy. She flies from Rome to Bangkok both on her way in and out and takes buses or rents a car to travel subsequently through Thailand, Vietnam, Laos, and Cambodia, before returning to Thailand to take her flight back home. According to the original UNWTO tourism data, there would be four trips: ITA→THA, ITA→VNM, ITA→LAO, and ITA→KHM. According to the GMP-revised tourism data [1], there would be eight trips: ITA→THA, THA→ITA, ITA→VNM, VNM→ITA, ITA→LAO, LAO→ITA, ITA→ KHM, and KHM→ITA. According to the air passenger data (regardless of distance-adjustment), there would be two trips: ITA→THA, THA→ITA. In reality, however, there were six trips: ITA→THA, THA→KHM, KHM→VNM, VNM→LAO, LAO→THA, and THA→ITA. In this case, both sources and all strategies lead to very different outcomes and none of them captures the transnational mobility that actually took place. This is an issue that has no easy solution. Structurally, it should lead to a slight overestimation of long-distance mobility between world regions (which is most likely when such round-trips are prone to occur). However, we argue that, compared to all global travels, this kind of journeys are rare and should not jeopardize the overall reliability of the dataset.

A second limitation is the following: by basing a substantial part of our mobility estimates on visitors who stayed overnight ('tourists' in the UNWTO terminology), we may be underestimating short-term border-crossings, for instance by commuters who live in border regions and regularly cross to the other side for work, leisure, or shopping. The following example is revealing in this regard: For the USA, detailed data on land border crossings are available (US Department of Transportation 2018). Looking at mobility between the USA and Canada, the distance-adjusted air passenger data [3] estimates about 20 million trips, while the GMP-revised tourism data [1] suggests around 33 million trips. The recorded land border crossing, by contrast, are 103 million—98 million private car passengers alone. Many of these moves are likely not overnight stays. While it is hard to generalize from this example,[13] it suggests that the mobility estimates in the Global Transnational Mobility Dataset (and the correction factor introduced in section 3.3)—although considerably larger than those provided by alternative global sources—are still quite conservative.

Finally, it is important to keep in mind that what the Global Transnational Mobility Dataset contains are mobility *estimates* rather than counts of actual, recorded trips. This is crucial. By applying a *statistical* approach to correct and adjust the data, we aimed at creating a revised dataset that *on average* captures mobility between countries more accurately. This procedure can however imply that in a minority of individual cases this revision leads to a more inaccurate estimate. We would thus like to remind that this dataset is well-suited to study structural features of transnational human mobility globally or for aggregates of countries. If the research interest is mobility between specific pairs of countries, the estimates in the Global Transnational Mobility Dataset are to be taken with caution, being aware of this limitation, and possibly comparing them to figures provided by alternative sources.

With these caveats in mind, we maintain that this novel dataset will prove to be a valuable resource for researchers interested in studying the global structure of transnational human mobility and its links to phenomena in the social and natural world, from wealth and well-being to the spread of epidemics and climate change.

[13] Table A1 in the Appendix provides an overview of the small number of cases in the UNWTO data where both 'visitors' and 'tourists' (i.e., overnight visitors) are reported. 'Tourists' as a share of 'visitors' range from 2 to 98 percent. The variance in this regard across countries is thus huge.

Appendix: Further details regarding the GMP revision of the UNWTO files

In section 3.1, we described the revision of the raw UNWTO data. In the following, several additional details regarding this process are given. First, several decisions had to be made to derive the preference order that is used when several categories of 'arrivals' are available in the same receiver country file (cf. Table 1 in the main text).

Issue 1: 'by nationality' vs 'by country of residence'

For almost all receiver countries, arrivals are reported *either* 'by nationality' *or* 'by country of residence'. In the few cases where both are reported,[14] we found that values do not differ dramatically between the two categories. If we were to decide for a restriction to one of these categories, we would lose a large percentage of cases (see category [1] vs [3] or [2] vs [4] in Figure A1). These two aspects taken together justify merging arrivals reported 'by nationality' and 'by country of residence' in a single dataset. In the rare cases where both categories are available, preference was given to 'by country of residence'.

Issue 2: 'tourists' vs 'visitors'

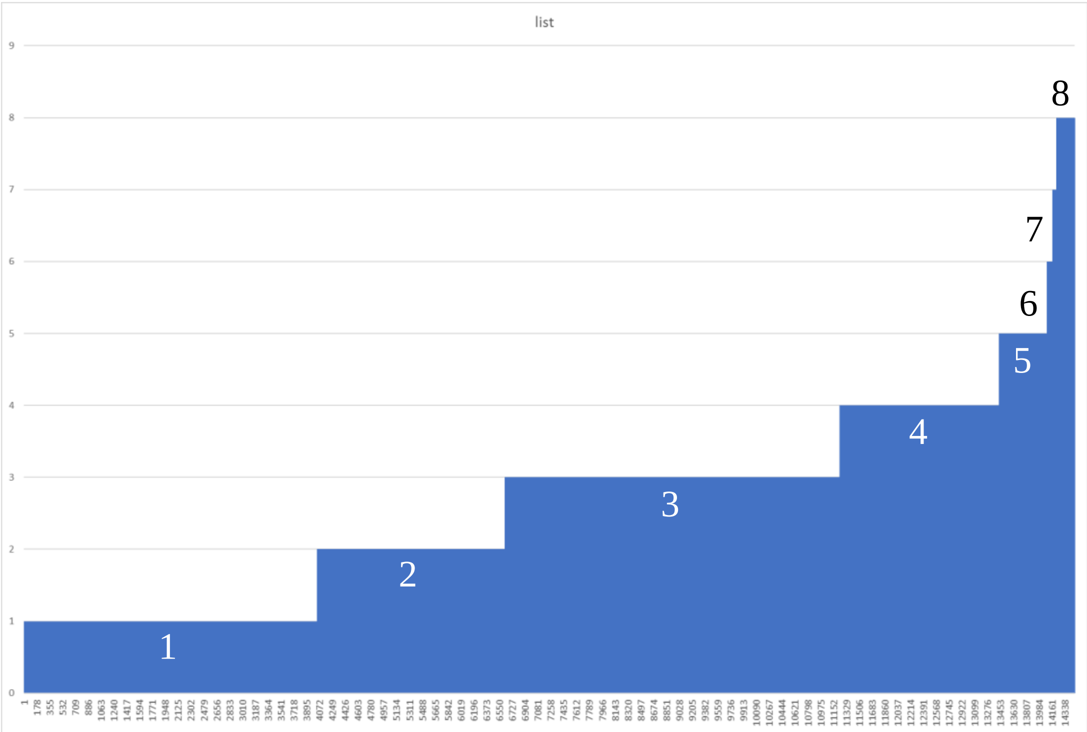
For a relative large percentage of cases, data on 'tourist' arrivals are unavailable and data on 'visitors' is reported instead (Categories [2] and [4] in Figure A1). We believe the benefit of not losing these cases outweighs the drawback of the imprecision that results from merging the two different categories in one dataset. According to the UNWTO definition (see section 3.1), 'visitor' is a broader category that includes 'both tourists and same-day non-resident visitors'. There are very few cases where country-to-country arrival data on *both* tourists *and* visitors is available (Table A1). However, in such cases, the size of the difference varies largely. In Venezuela, tourists as a share of visitors constituted 98 percent in 2010, whereas in Belarus it was only 2 percent, with the other thirteen countries being distributed quite evenly across the whole percentage range in between. (It seems plausible that in small countries the difference is more sizeable than in large countries). In the rare case that both 'tourists' and 'visitors' are reported we give preference to 'tourists' since the majority of cases are reported as tourists (categories [1], [3] and [5-8] in Figure A1, making it more or less the 'standard category').

Issue 3: 'at national borders' vs 'in accommodation establishments'

A third issue concerns the question of whether data collected 'at national borders' is comparable to data collected via 'accommodation establishments' (Table A2). To get an idea, we can draw on a total of 20 receiver countries for which both category types are available. In 17 out of these 20 countries, the number of arrivals at national borders is larger than the number of arrivals in accommodation establishments. A likely explanation is that some travelers who arrive in the country find private accommodation that is not covered in the data. In three exceptional cases (Iceland, Israel, Thailand), there are more arrivals in accommodation establishments than at national borders. On average, i.e., across all 460 cases (i.e., country-years) for which data is available, the ratio is .786, which could be interpreted as: on average, the number of arrivals reported for accommodation establishments is 78.6 percent the size of the number of arrivals reported at national borders. Note however, that the according standard deviation is very large (.456 or 46.5 percent) which makes the meaning and usability of this mean value questionable. The across-time variance within countries is much smaller (.085 or 8.5 percent on average), suggesting that individual countries are relatively consistent in their reporting style, while between countries there are considerable differences.

[14] These cases are, for 111 vs. 112: Guinea, Mali, Mexico, Nepal, Sri Lanka, and Thailand, for 121 vs 122: Indonesia, Macao, and Singapore. For 1911 vs. 1912 and 711 vs 712 no countries with both categories reported were found. In the cases of Guinea, Nepal, Mexico, Indonesia, and Macao, information was more detailed in the 111 and 121 categories than in the 112 and 122 categories.

Figure A1: Distribution of arrival categories in the 196-country version of the UNWTO dataset



Note: 1= Arrivals of non-resident tourists at national borders, by nationality; 2 = Arrivals of non-resident visitors at national borders, by nationality; 3 = Arrivals of non-resident tourists at national borders, by country of residence; 4 = Arrivals of non-resident visitors at national borders, by country of residence; 5 = Arrivals of non-resident tourists in all types of accommodation establishments, by country of residence; 6 = Arrivals of non-resident tourists in hotels and similar establishments, by nationality; 7 = Arrivals of non-resident tourists in hotels and similar establishments, by country of residence; 8 = Arrivals of non-resident tourists in all types of accommodation establishments, by nationality

Table A1. Tourists as a share of all visitors in 15 countries with both categories available, 2010

Country	Tourists	Visitors	Tourists as a share of visitors
Belarus	118,749	6,129,863	2%
Belize	241,919	1,197,326	20%
Hungary	9,511,000	39,905,000	24%
British Virgin Islands	330,343	842,497	39%
Jordan	4,207,408	8,078,380	52%
Hong Kong	20,085,155	36,030,331	56%
Italy	43,626,118	73,225,219	60%
Canada	16,219,399	25,621,300	63%
South Africa	8,073,552	11,303,087	71%
Israel	2,803,125	3,443,988	81%
Mongolia	456,963	557,452	82%
Namibia	984,098	1,114,423	88%
Saint Vincent and the Grenadines	72,478	77,564	93%
Turkey	31,364,004	32,997,308	95%
Venezuela	526,255	535,270	98%

Note: for Belarus, 2012 was used since 2010 was missing

Table A2. Arrivals 'in accommodation establishments' as a share of 'at national borders'

Categories of comparison/receiver country	Mean	SD across years
1911 as a share of 111 ('all types of accommodation', 'by nationality')		
Hungary	0.376	0.024
Iceland	2.412	0.175
Italy	0.975	0.078
Turkey	0.652	0.140
Cross-country mean	*1.104*	*0.105*
1912 as a share of 112 ('all types of accommodation', 'by country of residence')		
Cyprus	0.779	0.050
Philippines	0.985	0.004
Spain	0.716	0.117
France	0.530	0.027
Greece	0.549	0.051
Cross-country mean	*0.712*	*0.050*
711 as a share of 111 ('hotels, etc.', 'by nationality')		
Thailand	1.737	0.083
Hungary	0.342	0.024
Iceland	1.658	0.085
Italy	0.785	0.051
Turkey	0.647	0.142
Morocco	0.577	0.185
Tunisia	0.769	0.210
Chad	0.337	0.063
El Salvador	0.481	0.144
Bolivia	0.765	0.108
Cross-country mean	*0.810*	*0.109*
712 as a share of 112 ('hotels, etc.', 'by country of residence')		
Guinea	0.390	0.080
Mali	0.293	0.140
Cyprus	0.777	0.051
Philippines	0.374	0.055
Spain	0.587	0.080
France	0.427	0.033
Greece	0.536	0.051
Norway	0.876	0.035
Israel	1.114	0.165
Malta	0.773	0.007
Cross-country mean	*0.615*	*0.070*
Mean of all country-means	*0.776*	*0.085*
SD across all country-means		*0.469*
Global Mean/SD across all 460 country-years	*0.786*	*0.465*

Note: The underlying data stems from the whole time range, i.e., 1995 to 2016. Figures in red refer to countries in which exceptionally arrivals recorded in accomodation establishments are larger than those recorded at national borders

Table A3. Arrivals in 'hotels and similar establishments' as a share of arrivals in 'all kinds of accommodation establishments', 1995-2016

Categories of comparison/receiver country	Mean	SD across years
711 as a share of 1911 ('by nationality')		
Hungary	0.865	0.066
Iceland	0.694	0.039
Italy	0.811	0.021
Turkey	0.991	0.010
Czech Republic	0.904	0.048
Slovenia	0.776	0.045
Macedonia	0.908	0.036
Cross-country-mean	*0.850*	*0.038*
712 as a share of 1912 ('by country of residence')		
Cyprus	0.997	0.004
Philippines	0.380	0.055
Spain	0.828	0.068
France	0.749	0.030
Greece	0.977	0.005
Norway	0.650	0.022
Bulgaria	0.984	0.006
Croatia	0.453	0.041
Estonia	0.927	0.022
Poland	0.839	0.058
Denmark	0.340	0.161
Lithuania	0.867	0.029
Portugal	0.920	0.018
Romania	0.973	0.020
Sweden	0.623	0.036
Austria	0.732	0.005
Belgium	0.792	0.021
Germany	0.891	0.007
Luxembourg	0.732	0.062
Netherlands	0.777	0.030
Switzerland*	0.885	n.a.
Cross-country-mean	*0.777*	*0.035*
Mean/SD across all country-averages	*0.795*	*0.036*
SD across all country-means		*0.176*
Global Mean/SD across all 520 country-years	*0.794*	*0.178*

Note: *only available for one year

To get the most comprehensive picture possible, we use both categories but give preference to the category 'at national borders' wherever it is available. For the sake of consistency, we do the same in the exceptional cases of Iceland, Israel, and Thailand. There are 22 receiver countries for which *only* data on arrivals at accommodation establishments is reported (i.e., only [one/some of] the categories 711, 712, 1911, 1912 are available):

> Austria, Belgium, Bosnia & Herzegovina, Burkina Faso, Cape Verde, Croatia, Czech Republic, Denmark, Estonia, Germany, Lithuania, Macedonia, Norway, Luxembourg, Netherlands, Palestine, Portugal, Senegal, Slovakia, Slovenia, Switzerland, Togo.

In order not to lose these receiver countries, we keep them in the data, assigning preference to the categories as indicated in Table 1 in the main text. Given the calculations described above, it is possible that for these 22 countries arrivals are underestimated.

<u>Issue 4: 'all types of accommodation establishments' vs 'hotels and similar establishments'</u>

A fourth issue concerns the difference between 'all types of accommodation establishments' vs 'hotels and similar establishments'. Here, we can draw on 28 countries for which both category types are available to get an idea of the extent of the difference. Table A3 shows that, as one would expect, 'all kinds of accommodation establishments' is always the larger category. Across all 520 cases (i.e., country-years) for which we have data, arrivals in 'hotels and similar establishments' are on average 79.5 percent the size of arrivals in 'all types of accommodation establishments'. Note, however, that there is quite some variance between countries, with the share ranging from 34.0 percent in Denmark to 99.7 percent in Cyprus. The standard deviation across all country-years is .178 or 17.8 percent. To get the most comprehensive picture, we give preference to the category 'all types of accommodation establishments' whenever it is available.

Due to (a) the large variance between countries, which makes the average share rather meaningless and (b) the fact that most countries from which we could make inferences are European while most countries for which we lack information are African (which may result in deviating reporting styles), we refrain from using the information given in Table A.3 to create a factor to correct for the likely underestimation of the number of arrivals in five countries for which only arrivals in 'hotels and similar establishments' are reported. These countries are Burkina Faso, Cape Verde, Palestine, Senegal, and Togo. This implies that for these five receiver countries arrivals are likely underestimated.

<u>Issue 5: dealing with 'odd' travel origin categories</u>

Besides bringing order into the various 'arrival' categories, there are several 'odd' categories of origin of travels in the data that need to be dealt with. Their relative weights in the full dataset are shown in Table A4.

Table A4. 'Odd' categories of travel origin in the UNWTO data

Category	**Percentage**
1. Normal cases (e.g., 'Albania')	92.5
2. Country pairs (e.g., 'Canada, United States')	2.7
3. 'Nationals residing abroad'	1.0
4. 'USSR (former)'; 'Scandinavia'; 'Yugoslavia, SFR (former)'; 'Benelux' (6 cases)	0.01
5. 'Other countries of [world region, 'the world']'	2.9
6. 'All countries of [world region]'	0.9
All 'odd' categories	7.5
Lost arrivals after measures taken	approx. 3.8

Note: Percentage refers to the total number of tourist arrivals, not to the number of cases.

Category 1. Normal cases

The vast majority of cases (92.5 percent) are 'normal' cases, i.e., they state the number of arrivals from a specific sender country to a specific receiver country. They are thus in the appropriate format to be considered in a country-to-country flow matrix.

Category 2: Country pairs

In 45 cases, the sender is not an individual country, but one of seven country pairs:

> 'Australia, New Zealand'; 'Belgium / Luxembourg'; 'Canada, United States'; 'China + Hong Kong, China'; 'Czech Republic/Slovakia'; 'Serbia and Montenegro'; 'United Kingdom/Ireland'.

In order not to lose these cases (which include major sender country pairs such as 'Canada/United States', we split the number of arrivals reported for these cases into portions corresponding to the population size of the two sender countries in the according year weighted by the two countries' populations' general propensity to get involved in tourism. This general propensity to get involved in tourism is calculated from the overall number of arrivals from that country in all normal cases (i.e., Category 1).

Category 3: Nationals residing abroad

For 29 receiver countries, the sender category 'nationals residing abroad' is reported. These countries include:

> Algeria, Belize, Burkina Faso, Chile, Colombia, Congo DR, Cuba, Dominican Republic, Gambia, Grenada, Guinea, Iran, Jordan, South Korea, Mexico, Morocco, New Zealand, Nicaragua, Nigeria, Oman, Philippines, Rwanda, San Marino, Saudi Arabia, Togo, Tunisia, Turkey, Uruguay, and Yemen.

Since no clear country of origin (of the trip) can be identified for these cases, we decided to drop them.

Categories 4-6: Broad group of countries

Regarding categories 4-6, there are two main obstacles. First, the assumption that the tourists will be split according to the population distribution and their propensity to engage in tourism becomes rather questionable for such large groups of countries (think of 'other countries of the world'), and hard to compute. Furthermore, it would require determining, for each case in category 5-6, which countries were not listed from a specific world region since this varies from receiver country to receiver country. These efforts combined with the questionable quality of the outcome seem to justify neglecting these categories rather than imposing problematic assumptions about them. Accordingly, we drop and ignore these cases.

Following all the above-mentioned steps, the number of 'lost' arrivals (i.e., not imputable to any sending country) is reduced to 3.8 percent of all arrivals in the full original version of the dataset. It is important to note that these 3.8 percent of arrivals are likely to be not randomly distributed. Instead, most of them result from residual categories (e.g., 'Other countries in the world'). These residual categories are presumably often constructed when there are relatively few incoming visitors from distant parts the world. Thus, we assume that the lost cases are overwhelmingly long-distance travel.

To increase the comparability with the air passenger dataset, we excluded the following countries and territories:

> American Samoa, Anguilla, Aruba, Bonaire, British Indian Ocean Territory, Channel Islands, Christmas Island, Cocos (Keeling) Islands, Cook Islands, Curaçao, Democratic Yemen (former), Faeroe Islands, Falkland Islands (Malvinas), French Guiana, French Polynesia, Greenland, Grenada, Guadeloupe, Guam, Hawaii, Holy See, Isle of Man, Johnston Island, Liechtenstein, Martinique, Midway Islands, Montserrat, Netherlands Antilles, New Caledonia, Norfolk Island, Northern Mariana Islands, Pitcairn, Puerto Rico, Reunion, Saba, Saint Helena, Saint Pierre and Miquelon, Serbia, Sint Eustatius, Sint Maarten (Dutch part), South Sudan, Svalbard and Jan Mayen Islands,

Taiwan, Tokelau, United States Virgin Islands, Wake Island, Wallis and Futuna Island, Western Sahara.

What remains is a comprehensive set of 196 sender and receiver countries that also underlies the data used in Deutschmann (2016 and 2017).

Finally, as an overview for researchers interested in exploring the UNWTO tourism files more closely, we report the availability of categories of arrival by receiver country in Table A5.

Table A5. Categories of arrivals in the UNWTO dataset by receiving country

Country	111	121	112	122	1912	711	712	1911	1011	1012	2111	2112	Note
Afghanistan													
Albania		X											
Algeria		X											
Andorra			X										
Angola			X										
Antigua and Barbuda			X										
Argentina	X												
Armenia			X										
Australia				X									
Austria					X		X			X		X	
Azerbaijan				X			X			X			
Bahamas			X									X	
Bahrain		X											
Bangladesh	X												
Barbados			X										
Belarus	X	X											121 only since 2012
Belgium					X		X			X		X	
Belize	X	X											111 only since 1998
Benin			X										
Bermuda			X								X		
Bhutan	X								X				
Bolivia	X					X			X				111 only since 2006, 711 complete
Bosnia & Herzegovina					X							X	
Botswana			X										
Brazil			X										
British Virgin Islands			X	X									
Brunei Darussalam	X												
Bulgaria				X	X		X			X		X	
Burkina Faso						X			X				
Burundi	X												
Cambodia			X										
Cameroon		X				X			X				
Canada			X	X								X	
Cape Verde							X			X			

Cayman Islands			X										
Central African Republic	X												
Chad	X					X			X				
Chile	X												
China		X											
Colombia			X										121 empty
Comoros	X												
Congo DR	X												
Congo R				X			X			X			
Costa Rica	X												
Croatia					X		X			X		X	
Cuba				X									
Cyprus			X		X		X			X		X	
Czech Republic						X		X	X		X		
Denmark					X		X			X		X	
Djibouti													
Dominica			X										
Dominican Republic			X										
Ecuador		X											
Egypt		X							X				
El Salvador	X					X			X		X		
Equatorial Guinea													
Eritrea		X											
Estonia					X		X			X		X	
Ethiopia			X										
Fiji			X							X			
Finland				X	X					X		X	
France			X		X		X			X		X	
Gabon	X												
Gambia	X												
Georgia				X			X						
Germany					X		X			X		X	
Ghana	X												
Gibraltar													no file
Greece			X		X		X					X	
Guatemala				X									121 empty
Guinea	X		X				X						
Guinea-Bissau	X												
Guyana			X										
Haiti			X										
Honduras	X												
Hongkong			X	X									112 only since 1998
Hungary	X	X				X		X	X		X		
Iceland	X					X		X	X		X		
India	X												
Indonesia		X		X			X						
Iran		X											
Iraq		X											
Ireland			X									X	
Israel			X	X			X			X			
Italy	X	X				X		X	X		X		

Ivory Coast				X									112 empty
Jamaica			X									X	
Japan		X											
Jordan	X	X											
Kazakhstan				X									
Kenya				X						X			
Kiribati	X												
Kuwait		X											
Kyrgyzstan				X									112 empty
Laos		X											
Latvia				X	X							X	
Lebanon	X												
Lesotho				X									
Liberia													
Libya		X											
Lithuania					X		X			X		X	
Luxembourg					X		X			X		X	
Macao		X		X			X			X			
Macedonia						X		X	X		X		
Madagascar	X								X				
Malawi			X									X	
Malaysia	X						X						112 empty
Maldives	X												
Mali	X		X				X			X			711 , 1011 empty
Malta			X							X		X	112 empty
Marshall Islands			X									X	111 empty
Mauritania													
Mauritius			X									X	
Mexico	X		X										
Micronesia			X										
Moldova		X						X			X		
Mongolia	X	X											
Morocco	X					X			X				
Mozambique				X									
Myanmar	X								X				
Namibia	X	X											
Nauru													
Nepal	X		X										
Netherlands					X		X			X		X	
New Zealand				X									
Nicaragua	X												
Niger	X												
Nigeria		X											
Niue			X										
North Korea													
Norway					X		X			X		X	112 empty
Oman		X											
Pakistan	X												
Palau	X								X				
Palestinian						X			X				
Panama				X									
Papua New Guinea				X									112 empty
Paraguay	X												
Peru			X			X			X				

Philippines			X		X								712 empty
Poland		X			X		X			X		X	
Portugal					X		X			X		X	111,121,112,122 empty
Qatar			X										
Romania				X	X		X			X		X	
Russia		X											
Rwanda		X											
Saint Kitts and Nevis			X										
Saint Lucia			X										
Saint Vincent and the Grenadines			X	X									
Samoa				X									
San Marino		X											
Sao Tome and Principe	X												
Saudi Arabia	X												
Senegal						X			X				
Seychelles			X										
Sierra Leone			X							X			
Singapore		X		X									
Slovakia								X			X		
Slovenia						X		X	X		X		
Solomon Islands			X										
Somalia													
South Africa			X	X									
South Korea		X											
Spain			X		X		X			X		X	
Sri Lanka	X		X									X	
Sudan	X												
Suriname			X										
Swaziland				X			X						
Sweden				X	X		X			X		X	112 empty, 122 only since 2011
Switzerland					X		X			X		X	
Syria													several empty categories
Tajikistan				X									
Thailand	X		X			X			X				
TimorLeste			X										
Togo							X			X			
Tonga			X										
Trinidad and Tobago	X												112 empty
Tunisia	X					X			X				
Turkey	X	X				X		X	X		X		
Turkmenistan	X												
Turks and Caicos Islands			X										
Tuvalu	X												
Uganda			X										
Ukraine			X										
United Arab Emirates													Only empty categories

United Kingdom				X								X	
United Republic of Tanzania				X									
United States of America			X										
Uruguay		X											
Uzbekistan			X										
Vanuatu			X										
Venezuela	X	X											
Vietnam				X									
Yemen	X								X				
Zambia			X										
Zimbabwe				X									

Note: To decode the arrival category codes, cf. Table 1 in the main text.

Author contacts:

Ettore Recchi (corresponding author)

Migration Policy Centre

Robert Schuman Centre for Advanced Studies, EUI

Villa Malafrasca

Via Boccaccio 151

I-50133 Firenze (FI)

and

Sciences Po, Observatoire sociologique du changement (OSC), CNRS, Paris, France

Email: ettore.recchi@eui.eu

Emanuel Deutschmann

Migration Policy Centre

Robert Schuman Centre for Advanced Studies, EUI

Villa Malafrasca

Via Boccaccio 151

I-50133 Firenze (FI)

and

University of Göttingen, Institute of Sociology, Göttingen, Germany

Michele Vespe

European Commission

Joint Research Centre (JRC)

Via Enrico Fermi 2749

21027 Ispra (VA)

Italy

References

Abel, G.J., and N. Sander. 2014. Quantifying Global International Migration Flows. *Science* 343(6178): 1520-1522.

Centeno, M.A., M. Nag, T.S. Patterson, A. Shaver, and A.J. Windawi. 2015. The Emergence of Global Systemic Risk. *Annual Review of Sociology* 41: 65–85.

Deutschmann, E. 2016. The Spatial Structure of Transnational Human Activity. *Social Science Research* 59: 120–36.

Deutschmann, E. 2017. *Mapping the Transnational World: Towards a Comparative Sociology of Regional Integration*. PhD Dissertation, Jacobs University/University of Bremen.

Deutschmann, E., J. Delhey, M. Verbalyte, and A. Aplowski. 2018. The Power of Contact: Europe as a Network of Transnational Attachment. *European Journal of Political Research* 57(4): 963–88.

Duncalfe, L. 2018. ISO-3166 Country and Dependent Territories Lists with UN Regional Codes. Available at: https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes (last accessed 08/01/2019).

Fiorio, L., G. Abel, J. Cai, E. Zagheni, I. Weber, and G. Vinué. 2017. Using Twitter Data to Estimate the Relationship between Short-term Mobility and Long-term Migration. *Proceedings of the 2017 ACM on Web Science Conference*, 103-110.

Gabrielli, L., E. Deutschmann, F. Natale, E. Recchi & M. Vespe. 2019. Dissecting Global Air Traffic Data to Discern Different Types and Trends of Transnational Human Mobility. Unpublished manuscript.

Gerhards, J., S. Hans, and S. Carlson. 2017. *Social Class and Transnational Human Capital. How Middle and Upper Class Parents Prepare Their Children for Globalization*. London/New York: Routledge.

Hawelka, B., I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti. 2014. Geo-located Twitter as Proxy for Global Mobility Patterns. *Cartography and Geographic Information Science* 41(3): 260-271.

Helbling, M., and Teney, C. 2015. The Cosmopolitan Elite in Germany. Transnationalism and Postmaterialism. *Global Networks* 15(4): 446–468.

Kuhn, T. 2015. *Experiencing European Integration: Transnational Lives and European Identity*. Oxford: Oxford University Press.

Mau, S., J. Mewes, and A. Zimmermann. 2008. Cosmopolitan Attitudes through Transnational Social Practices? *Global Networks* 8(1): 1–24.

Mayer, T. and Zignago, S. 2006. GeoDist: The CEPII's Distances and Geo-graphical Database. MPRA Paper No. 31243.

Messias, J., F. Benevenuto, I. Weber, and E. Zagheni. 2016. From Migration Corridors to Clusters: The Value of Google+ Data for Migration Studies. *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 421-428.

Nye, J.S., Keohane, R.O. 1971. Transnational Relations and World Politics: An Introduction. *International Organization* 25(3): 329-349.

Rango, M., and M. Vespe. 2017. *Big Data and Alternative Data Sources on Migration: From Case-studies to Policy Support. Summary Report*. Ispra: Joint Research Centre of the European Commission.

Recchi, E. 2015. *Mobile Europe: The Theory and Practice of Free Movement in the EU*. Basingstoke: Palgrave Macmillan.

Recchi, E. 2017. Towards a Global Mobilities Database: Rationale and Challenges. Explanatory Note. MPC/EUI. Available at: http://www.migrationpolicycentre.eu /docs/GMP/Global_Mobilities_Project_Explanatory_Note.pdf (last accessed 3/3/2019).

Recchi, E., A. Favell, F. Apaydin, R. Barbulescu, M. Braun, I. Ciornei, N. Cunningham, J. Diez Medrano, D. N. Duru, L. Hanquinet, S. Pötzschke, D. Reimer, J. Salamonska, M. Savage, J. Solgaard Jensen, A. Varela. 2019. *Everyday Europe: Social Transnationalism in an Unsettled Continent*. Bristol: Policy Press.

Reyes, V. 2013. The Structure of Globalized Travel: A Relational Country-Pair Analysis. *International Journal of Comparative Sociology* 54(2): 144–70.

Sabre (2014). Aviation Data Intelligence, Leg Flow Tables (2014). Available at: http://www.sabreairlinesolutions.com/home/software_solutions/airports/ (last accessed 10/3/2019).

Spyratos, S., M. Vespe, F. Natale, I. Weber, E. Zagheni, and M. Rango. 2018. Migration data Using Social Media: A European Perspective. JRC Technical Report. doi: 10.2760/964282.

State, B., I. Weber, and E. Zagheni. 2013. Studying Inter-national Mobility through IP Geolocation. *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, 265-274.

United Nations High Commissioner for Refugees (UNHCR). 2016. Global Trends – Forced Displacement in 2016. Available at: https://www.unhcr.org/globaltrends2016/ (last accessed 9/1/2019).

United Nations World Tourism Organization (UNWTO). 2015. *Methodological Notes to the Tourism Statistics Database*. Madrid: UNWTO.

U.S. Department of Transportation. 2018. Border Crossing Entry Data. Available at: https://data.transportation.gov/Research-and-Statistics/Border-Crossing-Entry-Data/keg4-3bc2 (last accessed 9/1/2019).

Wimmer, A. and N. Glick Schiller. 2002. Methodological Nationalism and Beyond: Nation-state Building, Migration and the Social Sciences. *Global Networks* 2(4): 301-334.

World Bank. 2018. Population, Total. Available at: https://data.worldbank.org/indicator/SP.POP.TOTL (last accessed 9/1/2019).

Zagheni, E., I. Weber, and K. Gummadi. 2017. Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants. *Population and Development Review* 43(4): 721-734.