The Globalization of Refugee Flows

Xavier Devictor
Quy-Toan Do
Andrei A. Levchenko

WORLD BANK GROUP
Fragility, Conflict and Violence Global Theme
&
Development Economics
Development Research Group
April 2020
Abstract

This paper analyzes the spatial distribution of refugees over 1987–2017 and establishes several stylized facts about refugees today compared with past decades. (i) Refugees today travel longer distances. (ii) Refugees today are less likely to seek protection in a neighboring country. (iii) Refugees today are less geographically concentrated. And (iv) refugees today are more likely to reside in a high-income OECD country. The findings bring new evidence to the debate on refugee burden-sharing.
The Globalization of Refugee Flows

Xavier Devictor† Quy-Toan Do‡ Andrei A. Levchenko§

JEL codes: J15, F22, F55

Keywords: refugees, forced displacement, burden sharing

*We are grateful to Andy Foster, Çağlar Özdén, Philip Verwimp and workshop participants at the Research Conference on Forced Displacement in Copenhagen for helpful suggestions, and to Chau Hai Le for superb research assistance. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of the World Bank, its Board of Executive Directors, or the governments they represent.

†World Bank: xdevictor@worldbank.org
‡World Bank: qdo@worldbank.org
§University of Michigan, CEPR, and NBER: alev@umich.edu
1 Introduction

By the end of 2018, the world had witnessed its highest recorded number of forcibly
displaced people worldwide at 70.8 million persons, including 25.9 million who had
crossed a border and become refugees. While that number includes 3.5 million refugees
from older conflicts in Afghanistan and Somalia, it is also made up of around 10.5
million forcibly displaced persons from the recent crises in the Syrian Arab Republic,
Myanmar, and South Sudan.

The 1951 Convention Relating to the Status of Refugees (complemented by the 1967
Protocol) determines that refugee status shall be granted to any person who finds her
or himself displaced “owing to well-founded fear of being persecuted for reasons of race,
religion, nationality, membership of a particular social group or political opinion” (Art.
1.A.2.). Signatory states commit to provide treatment “no less favorable than nationals
of foreign countries in the same circumstances” with respect to employment (Art. 17),
housing (Art. 21), education, and public relief (Art. 22 and 23).

Most importantly, the Convention underlines the need for solidarity among countries
in sharing the responsibility of hosting refugees. Yet, the non-refoulement clause (Art.
33) implies that first countries of contact with asylum seekers are often those who have
to provide protection. Other signatories, on the other hand, can voluntarily decide on
their involvement in responsibility-sharing, potentially leading to free-riding (Suhrke
1998, Bubb, Kremer and Levine 2011). This creates a fundamental imbalance across
countries UN Member States, with political and fiscal constraints of host countries
having been associated with a lack of adequate assistance (Hathaway and Neve 1997,
Crisp 2003) and an additional reason for setting up refugee camps (Smith 2004).

On December 17, 2018, the United Nations General Assembly affirmed the Global
Compact on Refugees (UN General Assembly 2018), after two years of extensive consul-
tations led by the United Nations High Commissioner for Refugees (UNHCR) with UN
Member States, international organizations, refugees, civil society, the private sector,
and experts. The Global Compact on Refugees aims to provide a framework for more
predictable and equitable responsibility-sharing across countries. It opens by highlight-
ing that “there is an urgent need for more equitable sharing of the burden and respon-
sibility for hosting and supporting the world’s refugees, while taking account of existing
contributions and the differing capacities and resources among States.” It formally “in-
tends to provide a basis for predictable and equitable burden- and responsibility-sharing
among all United Nations Member States, together with other relevant stakeholders as
appropriate.” Underpinning the global debate on responsibility-sharing is the assump-
tion that “the grant of asylum may place unduly heavy burdens on certain countries”
(UN General Assembly 2018), typically countries neighboring a conflict area. In this
perspective, the number of refugees a country is to host is simply a function of its
geography.

In this paper, we examine empirically the proposition that the hosting of refugees
falls disproportionately on neighboring countries, which in most cases are in the devel-
oping world. To do so, we use data on worldwide bilateral refugee stocks compiled by
UNHCR to examine the spatial distribution of refugees and its evolution over time. Our
period of analysis is 1987-2017. We construct four outcome measures of refugee spatial
distribution. First, we compute the average distance refugees have traveled between
their country of origin and their country of destination. Next, we look at the probabil-
ity countries of origin and destination are contiguous. Third, we construct a measure
of refugee spatial dispersion by computing the Herfindahl index of refugee shares by
source country. Finally, to get some indication about where refugees go, we look at the
share of refugees seeking protection in high-income OECD countries.

To alleviate compositional issues, we project the matrix of distances traveled on
source country fixed effects and time effects. Source fixed effects control for time-
invariant country characteristics, and thus allows us to rule out that differences in
average distance traveled by refugees is driven by differences over time in which countries
experience conflict.

Our main findings can be summarized as follows. The average distance traveled
by refugees has increased substantially over time, and the share of refugees going to
an adjacent country has fallen. The Herfindahl index of refugee shares decreased sub-
stantially over time, indicating that refugees for a given source country are now more
dispersed across host countries. These results paint a picture of a more globalized and
far-reaching refugee network and imply a more equal distribution of the responsibility
of refugee hosting. In particular, we find that high-income OECD countries host an
increasing share of the refugee population. As of 1990, under 5 percent of refugees resided in a high-income OECD country. This share grew to nearly 25 percent by the mid-2000s, before falling somewhat to 15 percent, triple the 1990 value.

The theoretical literature on refugee hosting has advocated for an international system of quotas (Hathaway and Neve 1997), which could even be traded (Schuck 1997). However, there are few empirical analyses of refugee data that can inform policy. A notable exception is Dreher, Fuchs and Langlotz (2019), which looks at bilateral aid flows and argues that donor countries use aid as a way to reduce the flow of refugees entering their territory. As such, they establish the existence of some form of bargaining with transferable utility between potential host countries. To further the debate on refugee hosting, Bubb et al. (2011) discuss a system of financial transfers from richer countries to poorer ones for hosting refugees and at the same time distinguishing the international protection of asylum seekers from economic migration. One could see such mechanisms at work in recent cooperation agreements between the EU and third countries such as Jordan or Turkey (Temprano Arroyo 2019).

The rest of the paper is organized as follows. Section 2 describes the data used in the analysis. Section 3 presents the results. Section 4 concludes.

2 Data

Our analysis is primarily based on data on refugee stocks compiled by the UNHCR. UNHCR annually publishes the data on refugee stocks by source and destination country pair. The term “refugee” includes both refugees and asylum seekers. Under the 1951 Convention Relating to the Status of Refugees and the 1967 Protocol, a refugee is defined as “a person who has been forced to flee his or her country because of persecution for reasons of race, religion, nationality, political opinion or membership in a particular social group” (Art 1.A.2.).

The UNHCR Population Statistics Reference database contains data for the period 1951 – 2017 (released on June 19, 2019). The data set compiles annual stocks of refugees and asylum seekers at the source-destination level for 197 destination and 223 source countries. The ultimate source of the data is the authorities of each receiving country.
While in principle there are observations going back to 1951, coverage prior to the late 1980s is too sparse to be usable. Thus, our analysis covers the period 1987-2017. Overall, we have 112,522 non-zero observations for bilateral stocks over the period 1987-2017.

Since the data are not recorded at the individual level, we cannot reliably calculate refugee flows. Thus, the main variable used in the analysis is the refugee stocks. By definition, the stock of refugees in any particular year mixes individuals that arrived at different times. Since our main object of interest is changes in refugee behavior over time, analyzing stocks will if anything attenuate temporal differences.

To better approximate flows, we restrict the sample to large refugee events. A refugee event begins in the year in which the global stock of refugees from a particular source country first exceeds 25,000. An event ends when the stock falls below 25,000, or 10 years after initiation if the destination is an OECD country, whichever comes first. Capping the termination date of the event also puts earlier and later years in the sample on a more equal footing, as stocks in later years contain earlier vintages of refugees. An added benefit of restricting the sample to large refugee events is that this procedure also removes source countries with small numbers of refugees, that fled not due to armed conflict but for more idiosyncratic reasons. We check robustness of this approach in two ways: (i) using all of the refugee stock observations available in the data set, and (ii) computing refugee flows as the positive time differences in refugee stocks from year to year (setting negative time differences to zero). The results are robust to these two alternatives.

The data on bilateral distance and contiguity come from CEPII. The distance variable refers to the great circle distance between the most populated cities of each country in the pair. The contiguity indicator is equal to one if the two countries share a land border.

Figure 1 charts the global refugee population over time. The sharp increase in the number of refugees over the past decade is evident. Such refugee movements can impose significant impacts on the destination countries. Table 1 reports the top 10 destination and top 10 source countries in the most recent available year.

\[1\text{For OECD countries, UNHCR stops classifying an individual as a refugee if they have been in the country for more than 10 years.}\]
Figure 1: Global refugee population, 1987-2017

Note: This figure plots the global stock of refugees.

Table 1: Top 10 destination and top 10 source countries, 2017

<table>
<thead>
<tr>
<th>Destination</th>
<th>No. persons</th>
<th>Source</th>
<th>No. persons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turkey</td>
<td>3,789,119</td>
<td>Syrian Arab Rep.</td>
<td>6,455,328</td>
</tr>
<tr>
<td>Germany</td>
<td>1,399,554</td>
<td>Afghanistan</td>
<td>2,958,208</td>
</tr>
<tr>
<td>Pakistan</td>
<td>1,396,619</td>
<td>South Sudan</td>
<td>2,446,277</td>
</tr>
<tr>
<td>Uganda</td>
<td>1,395,115</td>
<td>Myanmar</td>
<td>1,197,794</td>
</tr>
<tr>
<td>Lebanon</td>
<td>1,014,165</td>
<td>Somalia</td>
<td>1,044,646</td>
</tr>
<tr>
<td>Iran (Islamic Rep. of)</td>
<td>979,519</td>
<td>Dem. Rep. of the Congo</td>
<td>756,899</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>932,319</td>
<td>Sudan</td>
<td>746,663</td>
</tr>
<tr>
<td>United States of America</td>
<td>929,762</td>
<td>Iraq</td>
<td>635,101</td>
</tr>
<tr>
<td>Sudan</td>
<td>924,789</td>
<td>Eritrea</td>
<td>564,447</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>891,990</td>
<td>Central African Rep.</td>
<td>558,487</td>
</tr>
</tbody>
</table>

Notes: This table reports the top 10 source and top 10 destination countries, measured by stocks of refugees recorded in 2017.
3 Main Result: Globalization of Refugee Movements

This section establishes the main result of the paper in three stages. We first display the unconditional trends. Second, we control for compositional differences. Third, we assess robustness of the main result to alternative data construction approaches.

**Unconditional trends** Figure 2 depicts the dimensions of the globalization of refugees. Panel (a) plots the average distance traveled by a refugee for each 5-year period between 1987 and 2017. There is a pronounced upward trend: the average distance traveled rises from around 1300km at the beginning of the sample to 2200km in the mid-2000s, before settling at around 1800km in the last decade. Panel (b) plots the share of refugees that find themselves in a country that shares a land border with their country of origin. At the beginning of the sample 95 percent of refugees were in a country contiguous with their home country. That share fell to 77 percent in the last decade.

Another manifestation of the increasing geographical reach of refugees is the greater number of destination countries to which they go. To document a more diversified set of destinations over time, we compute the Herfindahl index of refugee shares across destinations for each source country in each time period. That is, for a specific source country \( s \) and year \( t \), the Herfindahl index is defined as

\[
H_{st} = \frac{\sum_{d} \left( \frac{Stock_{sdt}}{\sum_{d'} Stock_{sd't}} \right)^2}{},
\]

where \( Stock_{sdt} \) is the number of refugees from \( s \) in \( d \) at time \( t \). The Herfindahl index takes a maximum value of 1 when all the refugees from \( s \) go to a single \( d \), so that \( d \)'s share is 1. The lower is the Herfindahl index, the more spread out is the pattern of refugee flows across destinations.

We then compute the simple mean of \( H_{st} \) for each year, and plot the 5-year averages of this mean. Figure 2, panel (c) reports the results. There is a substantial decrease in the Herfindahl of destination shares over time, from an average of 0.62 at the beginning to 0.36 at the end. The fall in the Herfindahl indicates greater diversification of refugee flows across locations over time.

---

2Because the Herfindahl is a simple average across source countries at a point in time, it may be dominated by smaller source countries that do not account for much of refugee flows. To check whether
Figure 2: Trends in refugee reach, 1987-2017

Note: This figure plots the average distance traveled by a refugee, the share of refugees finding themselves in a contiguous country, the average Herfindahl index of refugee shares by source country, and the share of refugees finding themselves in a wealthy OECD country.

Finally, to isolate the increasing impact of refugee inflows on high-income countries, panel (d) plots the share of the global refugee stock in high-income OECD country destinations\(^3\). The increasing importance of wealthy countries as hosts of refugees is evident.

Controlling for composition  Next, we assess whether the time trends documented above are driven by the changing composition of refugee source countries over time. For this is driving the results, we also examined the evolution of the average Herfindahls for only the top 10 and top 5 source countries in each year (which countries are in the top 10 or 5 changes from year to year, as different countries undergo conflicts). The pattern of increased source diversification is quite similar for the top source country sample. The results are available upon request.

\(^3\)High-income OECD countries in our sample are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and United States. We thus exclude the newer members of the OECD, such as the Republic of Korea, Mexico, or Turkey.
instance, if conflicts that give rise to refugee flows occurred in more remote countries in the more recent periods, then the distance traveled would increase. This would not be because it is now easier for refugees to travel farther, but rather because of the changing geography of conflict. To rule out pure compositional changes, we estimate the following regression at the source-time period level:

\[ \text{Outcome}_{st} = \delta_t + \delta_s + \varepsilon_{st}, \]

where \( \text{Outcome}_{st} \) is one of the four outcomes reported above – log average distance traveled by a refugee, share of refugees going to a contiguous country, the Herfindahl index of destinations, or share in a high-income OECD country – from country \( s \) in time period \( t \), and \( \delta_t \) and \( \delta_s \) are time and source country effects.

Source country effects imply that we are exploiting time variation within a source country over time in how far refugees travel. The coefficients of interest are the time effects \( \delta_t \). The regression is estimated weighted by total refugee stock, to obtain estimates of how outcome variables changed at a refugee, rather than country level. Standard errors are clustered at the source country level. Panel (a) of Figure 3 plots the time effects for the average distance traveled along with 95 percent confidence intervals. Since the distance traveled is in logs, the coefficients are interpretable as the percentage increase in the average distance traveled by a refugee in period \( t \) relative to the omitted period, which in our case is the first 5 years of data. The upward trend is evident, and the differences with respect to the initial period are statistically significant. In the final 5-year period, the average distance traveled is about 40 percent larger than in the first period. This proportional difference is quite similar in magnitude to the unconditional increase reported in Figure 2.

Panel (b) reports the time effects on the share of refugees found in a contiguous country. Since the left-hand side is a share, the time effect point estimates correspond to the difference in that share relative to the initial period. The share of refugees in a contiguous country falls by 16 percentage points after controlling for source effects. Once again, this difference is not far from the unconditional difference.

Panels (c) and (d) plot the time effects on the Herfindahl and the share going to high-income OECD countries. The trends evident in unconditional data are equally
strong when controlling for source country effects.

For all four outcome variables, the differences between the initial and later periods are highly statistically significant.

**Impact vs. diffusion over time** We next address the question of whether the trends documented in Figures 2-3 are due to the initial decision of refugees of where to flee from their homeland, or subsequent movements to third countries. Note that we cannot answer this question definitively without individual-level panel data. In our data, we do not observe the country from which a refugee entered their current host country, and thus cannot tell whether a given refugee in a given host country came from their homeland, or from yet another host country.

Nonetheless, we perform the following exercise. We are working with a set of refugee
Figure 4: Trends in refugee reach in event time

Note: This figure plots the four outcome variables in event time, for 4 sub-periods separately.

events defined in Section 2. An event is combination of a source country, a year of onset, and an end year. Thus, we can compute the evolution of all of our outcome variables – distance, contiguous share, Herfindahl, and share in high-income OECD – for each specific event and each year following its onset. We then plot these outcome variables in event time, with year 0 indicating the initial year of the event, up to year 10 of the event.

Figure 4 plots the four outcome variables in event time, for events starting in 4 different sub-periods. The main conclusion from this figure is that the differences across time in refugee reach are already apparent at the initiation of a refugee crisis. That is, distance traveled by a refugee rises monotonically from earlier to later refugee crises already in years 1 to 3 of a refugee crisis. While the pictures are somewhat noisy, it is not the case that the globalization of refugees trends documented above are due purely, or even primarily, to stronger diffusion of refugees over time.
Robustness Finally, we assess robustness of the results in a number of dimensions. First, since the beginning of our data, a number of country boundaries have changed. Prominent examples are the unification of Germany, the breakup of the Soviet Union and Yugoslavia, and the independence of South Sudan. To check whether the results are driven by changing country boundaries, Appendix Figure A1 reports the time effects conditional on source effects in a sample of entities where country definitions and boundaries are fixed throughout the sample period. There is virtually no difference in the results.

The second set of robustness checks probes our definition of refugee stocks. Appendix Figure A2 replicates the analysis using all refugee stocks available in the data, without constraining the sample to refugee events. The results are very similar to the baseline. Taking another approach, Appendix Figure A3 instead uses refugee flows. As argued above, without individual-level data, flows cannot be computed precisely. We build flows by taking annual time differences in stocks by source-destination pair. In some instances, stocks fall over time. Since we do not have confidence that a reduction in stocks represents a return to the home country – as opposed to transition to another host country – we set flows to zero whenever the difference in stocks is negative. As evidenced in the figure, the point estimates of the time effects and their statistical significance are quite similar for flows to the baseline.

Third, it may be that the destination-specific conditions (such as the global financial crisis) also affect the distance traveled by refugees, or the probability of not going to a contiguous country. To account for this possibility, we net out the time variation in the destination country conditions as follows.

In step 1, we project the refugee stocks at the source-destination-year level on source-time, destination-time, and source-destination fixed effects in a gravity-like specification:

\[ \text{Refugees}_{sdt} = \delta_{st} + \delta_{dt} + \delta_{sd} + \varepsilon_{sdt}. \]

We estimate this equation by Poisson Pseudo-Maximum Likelihood (Eaton, Kortum and Sotelo 2012), pooling countries and years (and thus including observations with zero bilateral stocks). We then construct a destination-adjusted refugee stock by subtracting
the destination-time effect from the actual stock:

\[ \text{AdjustedRefugees}_{sdt} = \text{Refugees}_{sdt} - \delta_{dt}. \]

Then, we compute the average distance traveled, share of refugees going to a contiguous country, the Herfindahl index of destinations, and share in wealthy OECD countries using this adjusted refugee data set instead of the actual data. Appendix Figure A4 reports the results. Netting out destination-time effects prior to carrying out the analysis leaves the main results virtually unchanged.

4 Conclusion

Our analysis suggests that the assumption underpinning the debate on responsibility-sharing may need to be partly revisited. Countries neighboring a conflict do host a majority of refugees and are hence bearing a disproportionate part of the responsibility to provide asylum to those who are fleeing from violence and oppression. Yet, the share of refugees who move to further-away destinations, including OECD countries, has been growing over time. In other words, responsibilities are increasingly shared across countries.

As it explores the notion of responsibility-sharing, the challenge for the international community is hence to determine how such trends can be sustained, at a pace which is optimal from a protection perspective, but also taking into account economic and political considerations across all potential refugee-hosting countries.

In parallel, it is important to recognize that the current responsibility-sharing remains deeply uneven. This is especially problematic as most refugee-producing crises are protracted, implying that the composition of the “main host countries” remains somewhat stable over large periods of time. Refugee burden sharing also implies, therefore, that increased support is warranted to maintain the current system and the international protection that it provides for those who are subject to persecution and violence.
References


Temprano Arroyo, Heliodoro, *Using EU aid to address the root causes of migration and refugee flows*, Florence: European University Institute, 2019.

Appendix
Figure A1: Trends in refugee reach, time effects controlling for source effects, harmonized country boundaries, 1987-2017

Note: This figure plots the time effects on the average distance traveled by a refugee, the share of refugees finding themselves in a contiguous country, the average Herfindahl index of refugee shares by source country, and the share of refugees finding themselves in a wealthy OECD country. Throughout, source country effects are netted out. The sample harmonizes all country boundaries across time.
Figure A2: Trends in refugee reach, time effects controlling for source effects, all stocks, 1987-2017

Note: This figure plots the time effects on the average distance traveled by a refugee, the share of refugees finding themselves in a contiguous country, the average Herfindahl index of refugee shares by source country, and the share of refugees finding themselves in a wealthy OECD country. Throughout, source country effects are netted out. The outcome variable is total stocks of refugees.
Figure A3: Trends in refugee reach, time effects controlling for source effects, flows, 1987-2017

Note: This figure plots the time effects on the average distance traveled by a refugee, the share of refugees finding themselves in a contiguous country, the average Herfindahl index of refugee shares by source country, and the share of refugees finding themselves in a wealthy OECD country. Throughout, source country effects are netted out. The outcome variable is total stocks of refugees.
Figure A4: Trends in refugee reach, time effects controlling for source effects, netting out destination-time effects, 1987-2017

Note: This figure plots the time effects on the average distance traveled by a refugee, the share of refugees finding themselves in a contiguous country, the average Herfindahl index of refugee shares by source country, and the share of refugees finding themselves in a wealthy OECD country. Throughout, source country effects are netted out. The analysis is carried out on an adjusted data set that nets out destination-time effects from every refugee stock observation.