



Regular Article

Refugees' and irregular migrants' self-selection into Europe[☆]Cevat Giray Aksoy^a, Panu Poutvaara^{b,*}^a European Bank for Reconstruction and Development, King's College London, IZA Institute of Labor Economics, UK^b Ifo Institute - Leibniz Institute for Economic Research at the University of Munich, University of Munich, CESifo, CReAM, IZA Institute of Labor Economics, Germany

ARTICLE INFO

JEL classification:

F22

J15

J16

J24

O15

Keywords:

International migration

Refugees

Irregular migrants

Self-selection

Human capital

Gender differences in migration

ABSTRACT

We provide the first large-scale evidence on self-selection of refugees and irregular migrants who arrived in Europe in 2015 or 2016. Our analysis uses unique datasets from the International Organization for Migration and Gallup World Polls. We find that refugees are positively self-selected with respect to human capital, as are female irregular migrants. Male irregular migrants are negatively self-selected. These patterns hold whether analyzing individually stated main reason to emigrate, country-level conflict intensity, or sub-regional conflict intensity. Several additional analyses show that our results are unlikely to be driven by omitted variable bias or liquidity constraints. We offer a theoretical framework to explain these patterns, by extending the Roy-Borjas model to include risks related to staying in an unsafe country of origin, risks related to migration, and gender-specific returns to human capital.

1. Introduction

The 1951 Refugee Convention and its extension in 1967 define a refugee as a person who is outside his or her country of nationality “owing to well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion,” and is unable or unwilling to return there.¹ An irregular

migrant, in contrast, is broadly defined as a person who travels abroad voluntarily in search of economic opportunities but has no legal right to remain in the intended destination country. The number of refugees and irregular migrants is at its highest level ever, even surpassing the post-World War II numbers (UNHCR, 2019). In 2016, the ten main source countries of refugees were in Asia and Africa, as well as nine of the ten largest refugee-hosting countries (see Fig. 1). Although most refugees

[☆] We are grateful to Nuno Nunes, Ivona Zakoska-Todorovska, and the International Organization for Migration (IOM) for kindly providing the Flow Monitoring Surveys. We also thank the editor, Andrew Foster, and two anonymous reviewers and Giacomo Battiston, Michal Burzynski, Daniele Coen-Pirani, Ralph De Haas, Anthony Edo, Guido Friebel, Kai Gehring, Yvonne Giesing, Thomas Ginn, Sergei Guriev, Hubert Jayet, Nadzeya Laurentsyeve, Till Nikolka, Caglar Ozden, Helmut Rainer, Carla Rhode, Dominik Sachs, Orkun Saka as well as participants at seminars at CEPS, CERDI, Institute for Employment Research (IAB), the University of Luxembourg, and the University of Trento, and the AFD-World Bank Migration and Development Conference, CEPR-LSE Migration Workshop, CESifo Conference on Employment and Social Protection, CReAM/RWI Workshop on Migration, COMPAS and Queen Mary Forced Displacement Conference, EPCS, German Economic Association Conference, IZA Annual Migration Meeting, OECD Migration Conference, SOLE, Silaplana Political Economy Workshop, University of Glasgow Migration and Mobility Workshop, and University of Pittsburgh Migration Workshop for their helpful comments. Zalina Alborova, Benedikt Behringer, Stefan Pasch, and Michael Ganslmeier provided outstanding research assistance. Views presented are those of the authors and not necessarily of the EBRD, IOM, or any other organization. All interpretations, errors, and omissions are our own.

* Corresponding author.

E-mail addresses: aksoyc@ebrd.com (C.G. Aksoy), poutvaara@ifo.de (P. Poutvaara).

¹ The Convention forbids returning a refugee to a territory where his or her life or freedom would be threatened due to persecution (UNHCR, 1967). In addition to those fleeing persecution, protection is typically granted also to those fleeing conflict. Distinguishing between refugees and irregular migrants is complicated: an irregular migrant has a strong incentive to claim to be a refugee to obtain permission to stay. Receiving countries, instead, may aim to reduce the number of refugees by tightening the burden of proof for persecution.

stay in developing countries (Chin and Cortes 2015; Hatton 2016, 2020), about 1.6 million refugees and irregular migrants arrived in Europe between January 2015 and December 2017 (see Fig. 2).²

Understanding the motivations of migrants, their self-selection in terms of socio-demographic characteristics, and their intended destination countries is important for several reasons. First, knowing the motivations of migrants helps to distinguish between the refugee crisis and challenges associated with irregular migration.³ Second, refugees' self-selection has implications for rebuilding their home countries. The more skilled refugees are, the more difficult it is to fill the gap they leave once the country enters the reconstruction stage. This loss is especially acute for poor countries, from which a significant share of refugees originates. Third, knowing the skill distribution and intended destinations of refugees who make it to the transit countries is helpful in planning integration policies, and thereby contributes to social stability in host countries and intended destination countries.⁴ Recent research shows that the refugee crisis has already had important political consequences in transit and intended destination countries.⁵

We make three main contributions in this paper. We start by presenting a theoretical explanation of how refugee self-selection in the presence of risks related to staying in an unsafe country and risks related to unsafe passage differs from the self-selection of irregular economic migrants. We then provide evidence on the motivations and intended destinations of both refugees and irregular migrants who arrived in Europe in 2015 or 2016 using data from the International Organization for Migration's Flow Monitoring Surveys (FMS). As a last step, we test our theoretical predictions by combining FMS with Gallup World Polls (GWP) to understand how migrant groups are self-selected from the origin population in terms of education and predicted income, and how

this self-selection differs between refugees and irregular migrants.⁶

Building on Roy (1951), Borjas (1987) showed that if skills are sufficiently transferable across countries, the theory suggests that migrants from a less egalitarian to a more egalitarian country should come from the lower end of the skill distribution, while migrants from a more egalitarian to a less egalitarian country should come from the upper end of the skill distribution. We show that a severe conflict can reverse the self-selection pattern from an inegalitarian country. The key mechanism is that a sufficiently high risk of staying in an unsafe country lowers expected returns to human capital below that available in more egalitarian destination countries. This novel result contrasts with the previous assumption that self-selection in terms of productivity is less intense for migrants who leave their home country based on non-economic motives (Chiswick, 1999, p. 181). It also differs from "refugee sorting" that Borjas (1987) proposed, based on the idea that skills that are rewarded in destination countries are penalized in origin countries. In our theoretical framework, human capital is rewarded in both origin and destination countries. Furthermore, we show that positive self-selection from countries plagued by major conflict can arise without assuming liquidity constraints. Liquidity constraints would further strengthen the result as those with more education are likely to have more resources that allow them to leave.

We find important differences in how refugees and irregular migrants are self-selected relative to the country-of-origin population in terms of their education. Both male and female migrants whose main reason to emigrate was conflict or persecution are positively self-selected with respect to secondary and tertiary education. Female irregular migrants (those motivated by other reasons than conflict or persecution) are also positively self-selected in terms of their education, although not as strongly as female refugees. Male irregular migrants, instead, are negatively self-selected in terms of their education. Negative self-selection of male irregular migrants is in line with Borjas (1987) prediction that migrants from countries with wider income differences should be negatively self-selected in terms of their productivity. Positive self-selection of women, instead, can be rationalized by severe gender discrimination in almost all origin countries as these imply a substantially lower rate of return for each year of education for women as compared to men. According to the United Nations' Gender Inequality Index (UN, 2017), all the origin countries we have in our sample score poorly and clearly below Western European countries.

We also analyze self-selection according to conflict intensity, irrespective of migrants' individually stated reason to emigrate. We compare how self-selection of migrants from major conflict countries with more than 1000 battle-related deaths (that is, major conflict countries) in a single year differs from self-selection of migrants from countries with fewer or no battle-related deaths (that is, minor or no conflict countries). Results are similar to those found using individually stated main reason to migrate: both male and female migrants from major conflict countries are positively self-selected in terms of their education. When it comes to migrants from minor or no conflict countries, women are positively self-selected, while negative self-selection pattern is found for men.

In our setting, some omitted factors, such as extractive institutions, could both increase the risk of conflict (see also Besley and Persson 2011) and depress opportunities available to those with higher education, pushing them to emigrate with a higher probability. This could generate a spurious correlation between self-selection at the upper part of the skill distribution and conflict. To address this concern, we exploit

² Refugee migration is also at the top of the political agenda in the United States, Canada, and Australia, which have resettled, respectively, about 3 million, 0.7 million and 0.5 million refugees since 1980 (Pew Research Center, 2019). Moreover, self-selection of irregular migrants is a much-debated topic in the United States. The estimated total number of irregular migrants in the United States was 10.5 million in 2017, representing 3.2 percent of the total U.S. population that year (Pew Research Center, 2019).

³ Irregular migrants do not have a legal right to stay in their intended destination country, but returning them to their country of origin is often costly and complicated. Irregular migrants often work in the shadow economy and pose a security concern as they may resort to crime, and those who have not applied for asylum have not been subject to normal security screening. They also face a higher risk of becoming a victim of crime, and their human rights are infringed in many transit and destination countries (UNHCR, 2008).

⁴ Two recent studies provide evidence on refugees' integration in host countries in the context of European refugee crises. Aksoy et al. (2020) study the causal effect of local labor market conditions and attitudes towards immigrants at the time of arrival on refugees' multi-dimensional integration outcomes (economic, linguistic, navigational, political, psychological, and social) in Germany. They find that high initial local unemployment negatively affects refugees' economic and social integration. They also show that favorable attitudes towards immigrants promote refugees' economic and social integration. Aksoy and Tumen (2021) show that good local governance practices can smooth out the refugee integration process and complement the efforts of central government in Turkey.

⁵ Hangartner et al. (2019) find that mere exposure to refugee arrivals induced a lasting increase in hostility towards refugees, immigrants, and the Muslim minority in Greek islands. Steinmayr (forthcoming) shows that the pattern is more complicated in Austria: exposure to refugees passing through border municipalities in 2015 increased far-right vote share, while contact and sustained interactions in municipalities hosting asylum applicants decreased far-right vote share. Ajzenman et al. (2020) document that anti-immigrant sentiment of natives increases considerably in localities that are more exposed to transit migration compared to those localities located further away.

⁶ We use the term "refugees and irregular migrants" to refer to individuals surveyed in Flow Monitoring Surveys. We will sometimes refer to each group separately. We will also sometimes refer to both groups as migrants. When analyzing self-selection with respect to education and predicted income, we restrict the attention to respondents aged 25 to 64, to focus on migrants who have likely already completed their education and are not yet retired.

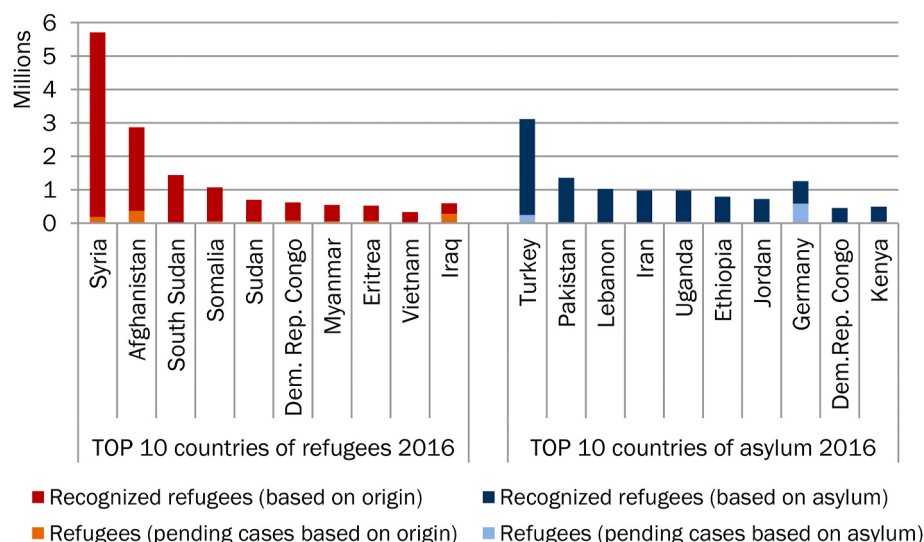


Fig. 1. Main countries of origin of refugees and main countries of asylum in 2016
Source: Eurostat (2017), UNHCR (2017) and authors' calculations.

variation in conflict intensity at the sub-regional level. We scrape geo-coded conflict data from Uppsala Conflict Database for all countries with major or minor conflict in the 2009–2014 period and use these data to identify high conflict and low conflict intensity sub-regions for each country, based on whether the number of battle-related deaths is above or below the country median. We then analyze how migrants from each sub-region are self-selected from the origin population living in that sub-region in terms of their education and other demographic characteristics.⁷ Consistent with our main results, we find that migrants who left from high conflict intensity sub-regions are positively self-selected in terms of education, as are female migrants from sub-regions with less intensive conflict, although to a lesser extent. Male migrants from sub-regions with less intensive conflict are, instead, negatively self-selected in terms of their education.⁸ We complement this analysis by using the methodology developed by Oster (2019) and find that our results are unlikely to be driven by unobserved variation: the estimates imply that for our results to be spuriously generated, unobserved factors must be 4 to 8 times as important as all observables and fixed effects included in the main specification.

Taken together, the self-selection patterns are similar when analyzing: (i) individually stated main reason to emigrate; (ii) country-level conflict intensity; and (iii) sub-regional conflict intensity. Additionally, we analyze self-selection among refugees and irregular migrants based on their main motivation to emigrate (that is, due to conflict vs. other reasons) in the migrant sample only and find consistent results with the rest of the analysis.⁹ To further investigate the self-

selection, we also compare the predicted earnings of refugees and irregular migrants and non-migrants. As GWP collects earnings at the household level, this analysis is restricted to singles who were in employment at the time of the interview. We find that refugees and female irregular migrants are strongly positively self-selected in terms of their predicted earnings, while male irregular migrants are negatively self-selected.

A major remaining challenge is how to distinguish the predictions of our theory from other potential explanations. Most importantly, one could be concerned that liquidity constraints alone would be able to explain the positive self-selection of migrants from major conflict countries as those with more education are more likely to be able to finance their trip. To address this concern, we directly control for the cost of journey between the origin country and the country of interview. We find that the cost of journey does not explain self-selection patterns in terms of education.¹⁰ Finally, we complement our analysis by investigating the self-selection patterns into Germany and Turkey.

The remainder of the paper is organized as follows: Section 2 reviews the related literature. Section 3 outlines a theoretical framework for understanding the self-selection mechanisms. Section 4 documents the data sources. Section 5 shows descriptive statistics. Section 6 describes the estimation strategy. Section 7 presents the results concerning self-selection of refugees and irregular migrants into Europe, after which section 8 concludes.

⁷ With the exception of Iran, each major or minor conflict country included in our sample was a colony (or part of a colony) before its independence. Country fixed effects capture the effects of colonial past.

⁸ We also find that main motivations of migrants are in line with our sub-regional conflict measure. In all countries, more migrants emigrated because of conflict or persecution from high conflict intensity sub-regions. For example, in Afghanistan, Iraq, and Syria, the share of men (women) emigrating from high conflict intensity sub-regions because of conflict or persecution is 50–60 (32–39) percentage points higher than among those emigrating from sub-regions with less intensive conflict.

⁹ This alleviates concerns that our results could be driven by the pseudo-cross-sectional sampling framework (i.e. migrant and source population samples coming from different surveys) we use in our main analysis. Migrants also did not have any incentive to lie as the survey was anonymous, no contact or other identifying information was collected, and the respondents were informed that their answers did not influence their legal status.

¹⁰ We also analyzed separately self-selection of migrants from poorest major conflict countries (Afghanistan, Nigeria, and Sudan). If liquidity constraints would be decisive in explaining the empirical patterns we observe, we should find positive self-selection from these poorest major conflict countries, independently of individually stated main reason to emigrate. However, we find results that are consistent with our theory: male and female refugees and female irregular migrants are positively self-selected with respect to their education, while male irregular migrants (who stated that their main reason to emigrate were economic reasons) are negatively self-selected. It is hard to see how negative self-selection of male irregular migrants could arise because of liquidity constraints.

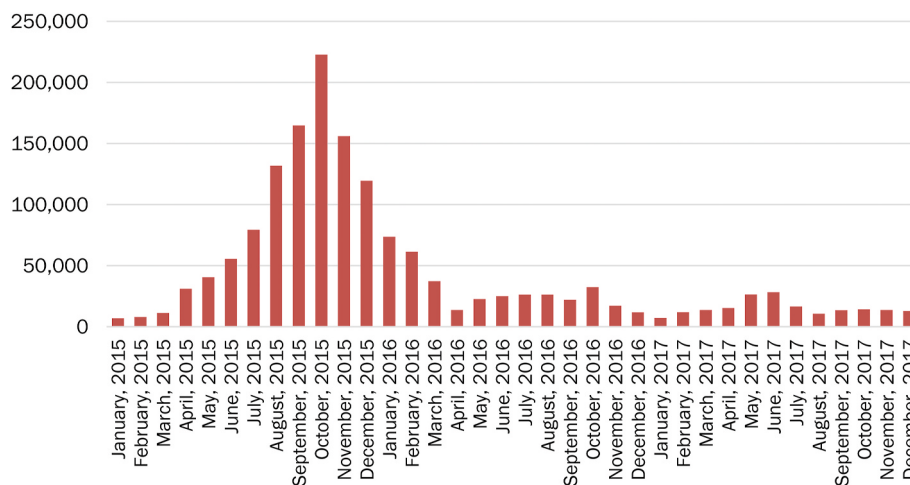


Fig. 2. Sea and land arrivals of refugees and irregular migrants to Europe, monthly.
Source: UNHCR (2018) and authors' calculations.

2. Related literature

Although there is a large literature on migrant self-selection, its focus has been on economic migrants.¹¹ There are only a few studies that analyze the self-selection of refugees from some individual countries. Cohen (2007) compares the educational levels and earnings assimilation of Jewish immigrants from the former Soviet Union in the United States and Israel and Birgier et al. (2018) examine the self-selection patterns in relation to Argentinian and Chilean refugees' economic assimilation in the United States, Sweden, and Israel. Brücker et al. (2016) show that recently arrived refugees in Germany have a relatively low level of education compared with the German population. Guichard (2020) analyzes asylum applicants who arrived in Germany from Afghanistan, Albania, Iraq, Serbia, and Syria. He finds that those from Afghanistan, Iraq, and Syria are more likely to have tertiary education than non-migrants, while those from Albania and Serbia are less likely to have tertiary education than non-migrants. These results are in line with what our theory predicts, but they could also be explained by differences in migration costs as traveling from Albania and Serbia is much cheaper and safer. In contrast, we analyze self-selection of male and female refugees and irregular migrants from multiple countries of origin into multiple destinations, in a setting in which both refugees and irregular migrants face comparable travel costs and risks, such as crossing the Mediterranean.

We also contribute to emerging literature on the effects of gender in migration decisions. Focusing on women's self-reported desire to migrate in Gallup World Polls, Ruyssen and Salomone (2018) concluded that "perceived gender discrimination forms a strong and highly robust incentive to migrate, but it does not seem to affect subsequent migration behavior." We extend the gender perspective on migration by analyzing the self-selection of refugees and irregular migrants separately by gender.

The European refugee crisis in 2015 and 2016 has had a much stronger political impact on Europe than the previous crises in the 1990s, as summarized by Dustmann et al. (2017). In 2015 and 2016, concerns about refugees and irregular migrants resulted in the re-introduction of border controls inside the Schengen area, disrupting the central principle of intra-EU free mobility and intra-European trade

and supply chains. Worries about immigration have powered the rise of populist parties and candidates in Austria (Halla et al., 2017), Denmark (Dustmann et al., 2019), France (Edo et al., 2019), Germany (Otto and Steinhardt 2014), Greece (Hangartner et al., 2019), and Italy (Barone et al., 2016). This makes it all the more important that we understand the socio-demographic characteristics of refugees and irregular migrants, many of whom are likely to apply for asylum in their destination countries.

Finally, our paper is also related to the literature on irregular migration. In an early theoretical contribution, Ethier (1986) shows that using a mixture of border controls and internal enforcement helps to reduce the costs of deterring illegal immigration. Friebe and Guriev (2006) present a model in which irregular migrants may enter into temporary servitude contracts to repay smuggler services and show that in that case, stricter deportation policies may backfire as they make enforcing servitude for irregular migrants in illegal sector easier. Most of the empirical research on irregular migration has focused on migration from Mexico to the United States (Hanson and Spilimbergo 1999; Orrenius and Zavodny 2005; Lessem 2017). We contribute to this literature by analyzing how irregular migrants arriving in Europe from multiple countries of origin are self-selected, and how this self-selection differs from the self-selection of refugees.

3. Theoretical framework

We analyze emigration from several countries of origin into one destination. We denote countries of origin with index k , and the potential destination with index d .¹² Our model builds on Borjas (1987) but adds into it gender-specific risks related to conflict or persecution in the home country and to migrating, and gender-specific returns to human capital. As Grogger and Hanson (2011), we simplify the Borjas (1987) framework by leaving out unobservable skill components in origin and in destination as our data do not allow testing hypotheses related to it. We denote the human capital of individual i beyond primary education by h_i . We denote gender by index g , $g \in \{f, m\}$, in which f denotes females

¹¹ See, for example, Borjas (1987); Hatton and Williamson (1998); Chiquiar and Hanson (2005); McKenzie et al. (2010); Fernandez-Huertas Moraga (2011); Grogger and Hanson (2011); Abramitzky et al. (2012); Belot and Hatton (2012); Parey et al. (2017); Hendricks and Schoellman (2018); Spitzer and Zimran (2018); Borjas et al. (2019).

¹² We model migration decision between staying in the home country or migrating into one destination, rather than a specific choice between various European countries, as refugees and irregular migrants may face considerable uncertainty on whether they can make it to their preferred destination country. In our empirical analysis, destination d may be interpreted as a composite of various European countries, in which different countries receive a weight corresponding to the conditional probability that a migrant from country k ends up there in case of reaching Europe.

and m denotes males. The wage individual i of gender g would receive in home country k in case of not being victim of conflict or persecution is given by

$$w_i^k = \exp(\alpha_{k,g} + r_{k,g}h_i),$$

in which $\exp(\alpha_{k,g})$ captures the wage available to those of gender g with primary education, and $r_{k,g}$ is the gender-specific return to human capital above primary education in country k . In case of a successful migration, individual i receives wage

$$w_i^d = \exp(\alpha_{d,g} + r_{d,g}h_i - \pi_{k,g}),$$

in which $\exp(\alpha_{d,g})$ depicts the wage available to natives of gender g with primary education in country d , $r_{d,g}$ is the gender-specific return to human capital above primary education in country d and $\pi_{k,g}$ measures the loss of potential productivity due to the imperfect applicability of migrant's human capital and lack of language skills. It plays an equivalent role to time-equivalent migration costs in Borjas (1987), and can differ across countries of origin, for example reflecting linguistic distance and quality of schooling.

The utility is logarithmic in terms of consumption in case of working and not becoming victim of conflict or persecution. To capture the risks associated with conflict or persecution, we assume that there is a country-specific and gender-specific risk $q_{k,g}$, $0 \leq q_{k,g} < 1$ of losing the utility from consumption and suffering an additional utility loss $L_{k,g}$, $L_{k,g} \geq 0$. It is reasonable to expect $q_{k,g}$ and $L_{k,g}$ to be highest in countries suffering war or civil war, intermediate in countries with a low-intensity conflict or persecution, while it can be viewed as zero, or close to zero, at least for men in relatively safe countries from which migrants are motivated by the lack of job opportunities, rather than conflict or persecution. For women, the risk term could also include gender-based violence and repression, like forced marriage, which could be a problem even in countries that are not suffering war or civil war. The expected utility in case of no migration is therefore

$$EU_i^k = (1 - q_{k,g})\log(w_i^k) - q_{k,g}L_{k,g}.$$

To capture the idea that refugees and irregular migrants face various risks also on their way to the destination countries, we assume that migrants of gender g from country k face risk $s_{k,g}$, $0 \leq s_{k,g} < 1$ of not making it to their intended destination in case of trying to migrate. This risk could include, at its extreme, dying on the way, as well as being caught in a transit country the migrant would prefer not to stay in because of imprisonment, lack of funds, or unexpected border closures, as well as being sent back to the country of origin in case asylum application is rejected. We denote the expected loss in case the risk is realized by $L_{M,g}, L_{M,g} > 0$. We do not take a normative stance on how high this loss should be perceived to be. From a positive perspective, the fact that thousands of people replied that they emigrated for economic reasons shows that the risk-adjusted loss was not prohibitively high for them. If the loss related to dying would be perceived to be infinitely high, we would not observe any irregular migration along the routes in which people die each year.

Individual-specific migration cost is denoted by ε_i . We assume that ε_i is independently and identically distributed for all individuals of the same gender in the same country. We denote its country-specific and gender-specific density function by $\phi_{k,g}$ and its cumulative distribution function by $\Phi_{k,g}$, which we assume to be differentiable. We assume that the distribution is sufficiently wide so that under all relevant circumstances, some people of each gender migrate and some stay. At the time of deciding whether to migrate, individuals know all the parameters of the model, including individual-specific migration cost and risk probabilities $q_{k,g}$ and $s_{k,g}$. The realizations of risks related to staying in an unsafe country of origin and to migration are revealed only after the decision on whether to migrate is made. The expected utility in case of migration is therefore

$$EU_i^d = (1 - s_{k,g})\log(w_i^d) - s_{k,g}L_{M,g} - \varepsilon_i.$$

It is ex-ante rational to migrate if $EU_i^d > EU_i^k$. This simplifies into the condition

$$\begin{aligned} \varepsilon_i < \varepsilon_i^* = & [(1 - s_{k,g})r_{d,g} - (1 - q_{k,g})r_{k,g}]h_i + (1 - s_{k,g})\alpha_{d,g} - (1 - q_{k,g})\alpha_{k,g} \\ & - (1 - s_{k,g})\pi_{k,g} + q_{k,g}L_{k,g} - s_{k,g}L_{M,g}. \end{aligned} \quad (1)$$

The probability that individual i migrates is given by

$$p_i = \Phi_{k,g}(\varepsilon_i^*). \quad (2)$$

Our main testable prediction relates to migrant self-selection:

Proposition 1. *Migrants are positively self-selected in terms of their human capital if $(1 - s_{k,g})r_{d,g} > (1 - q_{k,g})r_{k,g}$ and negatively self-selected in terms of their human capital if $(1 - s_{k,g})r_{d,g} < (1 - q_{k,g})r_{k,g}$.*

Proof. Inserting ε_i^* from equation (1) into equation (2) and partially differentiating gives $\frac{\partial p_i}{\partial h_i} = [(1 - s_{k,g})r_{d,g} - (1 - q_{k,g})r_{k,g}]\phi_{k,g}(\varepsilon_i^*)$.

Proposition 1 shows that the migrants are positively self-selected if $(1 - s_{k,g})r_{d,g} > (1 - q_{k,g})r_{k,g}$ and negatively self-selected if $(1 - s_{k,g})r_{d,g} < (1 - q_{k,g})r_{k,g}$. This condition refers to gender-specific risk-adjusted expected returns to human capital and it implies that if returns to skills are higher in the country of origin and the country of origin is relatively safe (a low risk term $q_{k,g}$) migrants are negatively self-selected, in line with Borjas (1987). However, if the country of origin suffers from a sufficiently severe conflict, given by $q_{k,g} > 1 - \frac{(1 - s_{k,g})r_{d,g}}{r_{k,g}}$, the self-selection is reversed, and migrants tend to come from the upper part of the skill distribution. If gender-specific return to human capital is lower in the country of origin than the risk-adjusted return $(1 - s_{k,g})r_{d,g}$ in the potential destination, migrants are positively self-selected even in the absence of risks of staying in the origin country.

Our model allows self-selection to differ between men and women. Given that gender discrimination is much more severe in African and Asian countries than in European destination countries, it is plausible to expect that even though average returns to human capital among men are higher in origin countries than in Europe, the reverse is the case for women. Therefore, we expect that the self-selection of men is negative from countries with low risk of conflict or persecution and positive from countries with a high risk of conflict or persecution, while self-selection of women should be positive from both low-risk and high-risk countries. Furthermore, we expect that the negative self-selection of male irregular migrants is stronger from more inequalitarian countries of origin.

Refugee self-selection has been previously studied also by Borjas (1987) and Chin and Cortes (2015), but with opposite predictions in comparison with our theoretical framework. Refugee self-selection in Borjas (1987) refers to the particular case in which skills which are rewarded in the destination country are penalized in the country of origin. When motivating this, Borjas (1987) refers to "countries that have recently experienced a Communist takeover." We analyze the case in which human capital is valued both in the countries of origin and in the destination, introducing risks related to migration and staying. According to Chin and Cortes, "refugees will be less selected on characteristics associated with labor market success in the destination country compared to other migrants." According to our theoretical framework, refugees' self-selection depends on the relative risks faced in the country of origin and during the migration process. Even if everyone in the country of origin faced the same risk of losing one's job due to persecution, an increase in such a risk could result in stronger self-selection in terms of human capital. The different prediction arises as Chin and Cortes (2015) model persecution as a disamenity which does not influence the wage income in the country of origin, while we assume that people who are subject to persecution both lose utility from their wage income and are subject to an additional utility loss, for example, due to

maltreatment if being imprisoned.

Although our empirical application focuses on migration from African and Asian origin countries into Europe, our theoretical model can be applied to other contexts. For example, if the aim would be to analyze the self-selection of migrants from Latin America to the United States, risk terms $q_{k,g}$ and $L_{k,g}$ could capture risks related to gang violence and drug wars, as well as conflict or persecution in countries such as Nicaragua and Venezuela. In the historical context, our model would generally predict that refugees are positively self-selected in case the risk term is sufficiently high.

4. Data

The data used in this paper come from the Flow Monitoring Surveys (IOM), Gallup World Polls, IAB-BAMF-SOEP Survey of Refugees in Germany, the World Bank's World Development Indicators (WDI), and the Uppsala Conflict Data Program (UCDP). The level of analysis is the individual level, and the details on how the dataset was constructed are provided below.

4.1. Flow Monitoring Surveys

Our analysis is based on the Flow Monitoring Surveys (FMS) obtained from the International Organization for Migration (IOM). The FMS provides in-depth insight into the profile, motivations, experiences, and intentions of migrants. It aims to derive quantitative estimates of the flow of (non-European) third-country nationals who are migrating towards Europe through the Central and the Eastern Mediterranean routes. The surveys were conducted in 11 languages (Arabic, Dari, English, Farsi, French, Italian, Kurdish, Pashto, Somali, Tigrinya, and Urdu) and administered by trained (male and female) data collectors with a mix of cultural and linguistic backgrounds (IOM, 2017). In the case of large groups, the surveys were conducted on a sample of migrants. In case of small groups, everyone was surveyed. FMS only gathers information from migrants aged 14 and older. To avoid repeated surveying of the same migrants, IOM field staff always ask whether respondents participated in the survey before starting the interview.

FMS provides rich information on migrants' country of origin, demographic characteristics (age, gender, educational attainment, and marital status), employment status before migration, reasons for leaving the place of residence, and intended destination. We use three waves of FMS in our main analyses. The first wave (October 2015 to December 2015) included interviews in Croatia, Greece, Slovenia, and the Former Yugoslav Republic of Macedonia.¹³ The second wave of FMS (January 2016 to November 2016) covers Bulgaria, Croatia, Greece, Hungary, the Former Yugoslav Republic of Macedonia, Serbia, and Slovenia. The third wave of the survey only covers Italy, spanning from June to November 2016. In the raw data, the three waves of FMS included 21,497 respondents. Furthermore, we have data for 13,469 respondents who were interviewed in Turkey as the fourth wave from November 2016 to August 2018. We report corresponding summary statistics for observations with non-missing responses to age, gender, and educational attainment questions in Table 1a and in the Online Appendix Table A.2. We focus on the first three waves in our main analysis for two reasons: (i) our main focus is the European refugee crisis; (ii) waves 1–3 were collected in 2015 and 2016 when the refugee crisis related to the Syrian Civil War peaked, while wave 4 was collected later.

The surveys aim to be representative of migrants arriving in Europe (IOM, 2017). The sample is large enough to gather evidence about migrants with different demographic profiles, and it is broadly reflective of

the migrant population (IOM, 2017). Online Appendix Fig. A.1 illustrates the main migratory routes across the Mediterranean: (i) the Central Mediterranean Route refers to the sea journey from sub-Saharan Africa to Italy, with Libya being the main point of departure; (ii) the Eastern Mediterranean Route refers to the sea crossing from Turkey to Greece; (iii) the Western Mediterranean Route refers to sea crossing from Morocco to mainland Spain, which is not part of the analysis. More than 99 percent of migrants who arrived in Europe in 2015 and 2016 using the Mediterranean routes did so through the routes we analyze. More specifically, 1,030,173 migrants arrived in Europe using the Eastern Mediterranean Route; 335,278 migrants used the Central Mediterranean Route; 13,400 migrants used the Western Mediterranean Route (UNHCR, 2017).

The surveys are conducted at the transit points.¹⁴ They are fully anonymous and voluntary. Respondents are approached by IOM field staff and are informed of the purposes of the research and of the fact that participation does not influence their legal status in the country of the interview. The rest of the questions are posed only to those migrants who give their consent to proceed. IOM also often uses data collector from the same nationality as migrants or at least speaking their language. The response rates were very high, which alleviates concerns about selection bias in participating in the surveys.¹⁵ Of those approached in the first two waves, only 2 percent declined to participate.¹⁶ The share declining to participate was somewhat higher in the third wave (4 percent) and fourth wave (5 percent).¹⁷

It is important to note that, despite the fast-changing conditions in the field, FMS provides a good representation of migrant groups. More specifically, in Online Appendix Table A.1, we evaluate the representativeness of FMS data from the first three waves by comparing it with official Eurostat data on asylum applications in 2015 and 2016. The table includes only those countries that were named as a country of nationality by at least 100 respondents in the FMS data. Most origin countries in our sample are predominantly Muslim and low or lower-middle-income nations. Overall, the Pearson correlation between the shares of the nationalities in FMS and Eurostat data is remarkably high: 0.98. The share of males is somewhat higher in FMS data, but the differences are small for most countries. The higher share of males in FMS than in Eurostat data may be partly explained by higher share of male respondents in Table 1a reporting to have come for economic reasons; such migrants may refrain from registering and, therefore, not show up

¹⁴ In Greece, all surveyed locations were on the mainland and not on the islands and represent a mix of official sites or camps managed by the national authorities and IOM premises in Athens. In the case of Balkan countries, the interviews were conducted both in spontaneous and organized transit points, including in official reception centers run by national authorities (IOM, 2017). As for Italy, the FMS was conducted in 36 entry and transit points in Sicily, Calabria, and Apulia in 2016 including 3 of the 4 hotspots operating in the country (Trapani, Pozzallo and Taranto). The hotspots are first reception facilities to identify and register migrants soon upon arrival (IOM, 2016).

¹⁵ There were no incentives provided for participation. It was on a voluntary basis and migrants could terminate the interview at any time. However, data collectors provided a clear explanation of the purpose of the survey before asking for consent and emphasized the fact that it was anonymous. IOM also selected locations for the survey where migrants stay longer and that provide the right environment in terms of space and confidentiality. These helped migrants to feel comfortable and respond to the survey.

¹⁶ The share of interviews by survey country in the first three waves is as follows: 24 percent in Italy, 20 percent in Greece, 19 percent in FYR Macedonia, 18 percent in Croatia, 6 percent in Bulgaria, 5 percent in Hungary, 5 percent in Slovenia, and 3 percent in Serbia.

¹⁷ In the first two waves (Eastern Mediterranean Route), interviewers approached 15,016 migrants on transit and only 286 of them did not participate the survey; in the third wave (Central Mediterranean Route), 7028 migrants were approached and only 261 of them gave no consent; and in the last wave (Turkey) only 730 migrants out of 14,199 did not want to take part in the survey. The overall response rate in the first three survey waves was 97.5%.

¹³ The Former Yugoslav Republic of Macedonia has officially adopted its new name "the Republic of North Macedonia" in February 2019. However, throughout the paper we use the Former Yugoslav Republic of Macedonia or FYR Macedonia, which were originally used in the surveys.

Table 1.a

Descriptive characteristics from Flow Monitoring Surveys Waves 1–3, full sample with no restrictions (age 14 and older).

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Full sample	Males	Females	Wave 1 (10/2015 - 12/2015)	Wave 2 (01/2016 - 11/2016)	Wave 3 (06/2016 - 11/2016)
Age	26.41 (8.43)	26.09 (8.24)	27.93 (9.15)	27.40 (8.83)	27.77 (8.71)	22.55 (6.01)
Male	0.82 (0.38)	–	–	0.79 (0.40)	0.81 (0.39)	0.88 (0.33)
Secondary education	0.50 (0.50)	0.50 (0.50)	0.48 (0.50)	0.40 (0.49)	0.53 (0.50)	0.46 (0.50)
Tertiary education	0.18 (0.39)	0.17 (0.38)	0.22 (0.41)	0.36 (0.48)	0.20 (0.40)	0.07 (0.25)
Employed	0.47 (0.50)	0.51 (0.50)	0.25 (0.43)	–	0.47 (0.50)	0.47 (0.50)
Married	0.30 (0.46)	0.25 (0.43)	0.55 (0.50)	–	0.42 (0.49)	0.17 (0.38)
Divorced	0.01 (0.01)	0.00 (0.06)	0.03 (0.18)	–	0.01 (0.11)	0.00 (0.08)
Widowed	0.01 (0.01)	0.00 (0.06)	0.03 (0.18)	–	0.01 (0.10)	0.00 (0.07)
<i>Main reason to emigrate:</i>						
Conflict or persecution	0.77 (0.42)	0.76 (0.43)	0.84 (0.36)	0.86 (0.35)	0.79 (0.40)	0.68 (0.47)
Economic reasons	0.18 (0.38)	0.19 (0.40)	0.09 (0.29)	0.11 (0.31)	0.17 (0.38)	0.21 (0.41)
Limited access to amenities	0.02 (0.15)	0.02 (0.15)	0.02 (0.14)	0.01 (0.08)	0.01 (0.07)	0.07 (0.26)
Other reasons	0.02 (0.15)	0.02 (0.14)	0.03 (0.18)	0.02 (0.10)	0.02 (0.14)	0.03 (0.17)
<i>Nationalities:</i>						
Syria	0.28 (0.45)	0.25 (0.43)	0.46 (0.50)	0.45 (0.50)	0.36 (0.48)	0.01 (0.11)
Afghanistan	0.20 (0.40)	0.22 (0.41)	0.16 (0.36)	0.23 (0.42)	0.28 (0.45)	0.01 (0.09)
Iraq	0.10 (0.30)	0.10 (0.30)	0.12 (0.32)	0.14 (0.34)	0.14 (0.34)	0.01 (0.09)
Nigeria	0.06 (0.25)	0.06 (0.25)	0.10 (0.30)	0.01 (0.07)	0.00 (0.05)	0.22 (0.41)
Pakistan	0.05 (0.22)	0.06 (0.25)	0.00 (0.07)	0.06 (0.23)	0.07 (0.25)	0.02 (0.13)
Eritrea	0.04 (0.21)	0.05 (0.23)	0.05 (0.10)	0.01 (0.04)	0.00 (0.03)	0.14 (0.35)
Morocco	0.04 (0.20)	0.05 (0.21)	0.02 (0.15)	0.02 (0.15)	0.05 (0.22)	0.03 (0.17)
Iran	0.04 (0.20)	0.03 (0.18)	0.05 (0.21)	0.05 (0.22)	0.05 (0.21)	0.01 (0.05)
Guinea	0.03 (0.18)	0.03 (0.17)	0.01 (0.09)	0.01 (0.02)	0.01 (0.03)	0.10 (0.31)
Bangladesh	0.02 (0.14)	0.03 (0.16)	0.00 (0.02)	0.01 (0.08)	0.00 (0.06)	0.07 (0.25)
N	20,465	16,781	3684	1667	13,063	5735

Notes: Means (standard deviations). The sample sizes for some variables are different either due to missing data or because they were not asked in each wave. Secondary and tertiary education refer to the highest completed education. We report shares for the top 10 nationalities in the full sample. Source: Flow Monitoring Surveys, 2015 and 2016 (waves 1–3).

in Eurostat data on asylum applications.

4.2. IAB-BAMF-SOEP Survey of Refugees

To investigate self-selection into Germany as the main destination country in Europe, we use the IAB-BAMF-SOEP Survey of Refugees (2016), which is collected as part of the German Socio-Economic Panel. The dataset provides information on refugees' country of origin, demographic characteristics, and employment histories and covers about 4000 adult refugees in Germany. Due to low number of observations on migrants from minor or no conflict countries and to make the analysis comparable with the rest of the paper, we only focus on refugees who arrived in Germany in 2015 and 2016 and migrated from major conflict countries with more than 100 respondents (Afghanistan, Iraq, Pakistan, Syria).

4.3. Gallup world poll and country level characteristics

Our primary data on the origin country population come from the 2009–2014 GWP.¹⁸ These nationally representative surveys are fielded every year in over 120 countries and interview approximately 1000 individuals in each country on a range of topics. The GWP provides detailed information on individuals' demographic characteristics (age, gender, educational attainment, and marital status), labor market outcomes, and income.

The manners in which education is reported vary by country, making cross-national comparisons difficult. To enable comparisons across

¹⁸ Specifically, we observe the data for Syria between 2009 and 2013, for Algeria between 2010 and 2014, for Iran between 2011 and 2014 and for Morocco between 2010 and 2013. For the rest of the source countries, we have data spanning from 2009 to 2014.

countries by educational attainment, GWP creates consistent categories (rather than providing continuous measure) and classifies all education descriptions within three categories: elementary (up to eight years of basic education), secondary (nine to 15 years of education), and tertiary (completed four years of education beyond “high school” and/or received a four-year college degree). Therefore, throughout the paper we use dummy variables to measure self-selection with respect to education. The GWP’s main advantage for our purposes is that the poll allows us to combine the FMS data with data on the country of origin (henceforth, non-migrant) population for a broad spectrum of countries.¹⁹ Specifically, we pool two unique datasets based on 13 origin countries reported by at least 100 respondents aged 25 to 64 in the first three waves of FMS.²⁰ We then create a “migrant” indicator variable, which takes a value of 1 for respondents surveyed in the FMS and zero otherwise. Using this pseudo-cross-sectional sample, we investigate how refugees and irregular migrants are self-selected from the origin population. Importantly, each variable in GWP is harmonized with the definitions used in FMS to ensure comparability.

To further understand the characteristics of refugees and economic migrants, we use Uppsala Conflict Data Program (UCDP) battle-related deaths dataset and classify the origin countries based on their conflict intensity (Pettersson and Eck, 2018).²¹ Following the definitions provided by UCDP, we define: i) *major conflict* category as countries with 1000 or more battle-related deaths in any of the years between 2009 and 2014 (this category includes Afghanistan, Iraq, Nigeria, Pakistan, Sudan, Syria); ii) *minor conflict* category as countries with 25–999 battle-related deaths in any of the years between 2009 and 2014 (this category includes Algeria and Iran); iii) *no conflict* category as countries that did not experience a major war or minor conflict in any of the years between 2009 and 2014 (this category includes Bangladesh, Cameroon, Côte d’Ivoire, Morocco, and Senegal). It is important to note that: (i) Political Terror Scale (Gibney et al., 2018) gives us the same classification using the data between 2009 and 2014; (ii) there is some movement of countries between the categories across years; (iii) using a continuous measure of the conflict intensity produces qualitatively similar results.²² We also use geo-coded conflict data from UCDP to classify each origin country’s sub-regions by their conflict intensity.

5. Descriptive statistics

Table 1a and 1b presents descriptive statistics of FMS data from Waves 1–3, with respect to demographic and socioeconomic characteristics of survey respondents. In Table 1a, we show the descriptive statistics for the full sample (i.e., with no age restrictions, and non-missing responses to age, gender, and educational attainment questions). It is also important to note that only about 4 percent of the respondents in the raw data has missing information on age, gender, and educational attainment questions.

In Table 1b, we report the characteristics of the sample of migrants aged 25 to 64. We focus on this age group when analyzing self-selection according to education and predicted earnings as those younger than 25 have often not yet completed their education, while those older than 64 may have already retired. We are not able to include Eritrea in our main analysis as it is not surveyed by GWP. We also exclude nationalities with less than 100 respondents. Our final sample in Table 1b consists of 9828

observations provided by migrants of 13 different nationalities (i.e., origin countries) with at least 100 respondents.

Table 1a shows descriptive statistics for the main FMS sample without age restrictions. An overwhelming majority of survey respondents are male (82 percent) with an average age of 26. 30 percent of the respondents are married, and 18 percent have tertiary education. About half of the respondents also report having been employed before migration. 77 percent cited conflict or persecution and 18 percent economic reasons as the main reason why they had emigrated. Limited access to basic services (like school and health care) or lack of food or accommodation was named as the main reason by only 2 percent of respondents. 2 percent referred to natural disasters or other reasons.

The respondents were also asked their intended destination country, which we report by survey waves. As shown in Online Appendix Fig. A.2, Germany was the main destination in the first two waves, chosen by 61 percent of respondents, followed by France (6 percent), Italy, and Sweden (5 percent each). In wave 3 that was collected in Italy, 61 percent reported an intention to stay in Italy, followed by 8 percent aiming to Germany and 6 percent aiming to France. In wave 4, 63 percent intended to stay in Turkey, 7 percent aimed to Germany, and 18 percent did not know where to aim. It is plausible that this reflects the effective closing of the route from Turkey to the European Union following the March 2016 agreement that allows the European Union to send migrants arriving from Turkey to Greece illegally back in exchange for taking recognized refugees from Turkey.

When we focus on individuals aged 25 to 64 in Table 1b, we again find that a vast majority of survey respondents are male (80 percent). There are also notable differences between the two samples: migrants in this age band are more likely than those in the full sample to be married (57 percent), to have tertiary education (25 percent), and to have been employed before migration (63 percent). These differences in education levels and employment status before migration also highlight the importance of focusing on individuals aged 25 and older when testing our theory. By doing so, we avoid the share of young respondents mechanically driving results for self-selection with respect to tertiary education and employment. Below, we also show that our results remain robust when we focus on those aged 14 and above.

Online Appendix Table A.3 shows descriptive statistics for the IAB-BAMF-SOEP Survey of Refugees. As this survey included only adults, it is not surprising that all respondents are somewhat older and more educated than in the FMS. If restricting the attention to those aged 25 to 64, the main difference between IAB-BAMF-SOEP Survey of Refugees and the FMS is that the share of women is clearly higher in Germany, and respondents are on average older.²³ In terms of education, differences compared with the average if pooling Waves 1 and 2 (in which Germany is the stated main destination) are small.

In Table 2, we present descriptive characteristics of the origin population from the GWP.²⁴ Unlike in Tables 1a and 1b, the gender ratio is balanced. People in origin countries (compared with refugees and irregular migrants) are also older on average, more likely to be married than those in the FMSs, and less likely to have (completed) tertiary education.

Online Appendix Table A.5 presents the differences in educational

¹⁹ Of course, it is possible that respondents surveyed in GWP might migrate later. This means that if the scale of emigration is large, we underestimate differences between migrants and non-migrants.

²⁰ The data come from the following thirteen countries: Afghanistan, Algeria, Bangladesh, Cameroon, Côte d’Ivoire, Iran, Iraq, Morocco, Nigeria, Pakistan, Senegal, Sudan, and Syria.

²¹ We use the UCDP’s best estimate for battle-related deaths to classify the countries based on their conflict intensity.

²² These results are not presented here but available upon request.

²³ Brücker et al. (2016) allowed multiple responses for reason to migrate and report that 70 percent of respondents listed war and conflict as reason to migrate, followed by persecution (44 percent), poor personal living conditions (39 percent), discrimination (38 percent), and fear of forced conscription (36 percent).

²⁴ Table 2 reports unweighted descriptive statistics from the Gallup World Polls. Importantly, we find that using Gallup sampling weights does not substantially change the shares we report in Table 2. These estimates are not reported here but available upon request. We also show the descriptive characteristics from the Gallup World Polls between 2009 and 2011 (i.e. pre-Syrian conflict period) in Online Appendix Table A.4.

Table 1.b

Descriptive characteristics from Flow Monitoring Surveys Waves 1–3, restricted sample with adults aged 25 to 64.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Full sample	Males	Females	Wave 1 (10/2015 - 12/2015)	Wave 2 (01/2016 - 11/2016)	Wave 3 (06/2016 - 11/2016)
Age	32.42 (7.60)	32.20 (7.39)	33.29 (8.36)	32.70 (8.32)	32.92 (7.80)	29.60 (5.03)
Male	0.80 (0.40)	–	–	0.80 (0.39)	0.79 (0.41)	0.87 (0.33)
Secondary education	0.48 (0.50)	0.49 (0.50)	0.45 (0.50)	0.35 (0.47)	0.50 (0.49)	0.44 (0.50)
Tertiary education	0.25 (0.43)	0.25 (0.43)	0.28 (0.45)	0.41 (0.49)	0.26 (0.44)	0.12 (0.32)
Employed	0.63 (0.48)	0.73 (0.44)	0.28 (0.45)	–	0.61 (0.49)	0.69 (0.46)
Married	0.57 (0.49)	0.52 (0.50)	0.78 (0.41)	–	0.65 (0.47)	0.40 (0.50)
Divorced	0.01 (0.13)	0.01 (0.10)	0.05 (0.21)	–	0.02 (0.14)	0.01 (0.12)
Widowed	0.01 (0.16)	0.00 (0.08)	0.05 (0.21)	–	0.01 (0.13)	0.01 (0.11)
<i>Main reason to emigrate:</i>						
Conflict or persecution	0.79 (0.41)	0.77 (0.42)	0.86 (0.35)	0.86 (0.34)	0.79 (0.41)	0.69 (0.46)
Economic reasons	0.17 (0.37)	0.19 (0.39)	0.09 (0.28)	0.10 (0.30)	0.17 (0.37)	0.22 (0.41)
Limited access to amenities	0.02 (0.13)	0.01 (0.11)	0.01 (0.10)	0.02 (0.08)	0.01 (0.07)	0.06 (0.23)
Other reasons	0.02 (0.16)	0.02 (0.14)	0.04 (0.19)	0.02 (0.08)	0.03 (0.16)	0.02 (0.15)
<i>Nationalities:</i>						
Syria	0.37 (0.48)	0.33 (0.47)	0.52 (0.50)	0.51 (0.50)	0.41 (0.49)	0.03 (0.07)
Afghanistan	0.17 (0.37)	0.18 (0.38)	0.15 (0.36)	0.15 (0.36)	0.20 (0.40)	0.01 (0.06)
Iraq	0.14 (0.35)	0.14 (0.35)	0.13 (0.34)	0.15 (0.36)	0.16 (0.37)	0.02 (0.15)
Nigeria	0.05 (0.21)	0.05 (0.22)	0.04 (0.20)	0.01 (0.08)	0.00 (0.05)	0.30 (0.46)
Pakistan	0.06 (0.23)	0.07 (0.25)	0.01 (0.08)	0.06 (0.24)	0.06 (0.23)	0.03 (0.17)
Morocco	0.05 (0.20)	0.05 (0.22)	0.03 (0.17)	0.02 (0.13)	0.05 (0.21)	0.18 (0.38)
Iran	0.06 (0.23)	0.05 (0.23)	0.07 (0.25)	0.07 (0.24)	0.06 (0.25)	0.01 (0.07)
Bangladesh	0.01 (0.12)	0.02 (0.13)	0.00 (0.02)	0.01 (0.07)	0.01 (0.05)	0.07 (0.26)
Senegal	0.01 (0.10)	0.01 (0.12)	0.00 (0.06)	0.01 (0.05)	0.01 (0.06)	0.06 (0.24)
Sudan	0.01 (0.11)	0.01 (0.09)	0.00 (0.07)	0.01 (0.06)	0.01 (0.05)	0.05 (0.20)
N	9828	7865	1963	925	7475	1428

Notes: Means (standard deviations). Respondents from Eritrea and countries with less than 100 respondents are excluded. The sample sizes for some variables are different either due to missing data or because they were not asked in each wave. Secondary and tertiary education refer to the highest completed education. Source: Flow Monitoring Surveys, 2015 and 2016 (waves 1–3).

attainment for the main origin countries. Apart from Nigeria, Bangladesh, and Senegal, migrants are better educated than non-migrants, whether analyzing the full population or restricting the attention to those aged 25 to 64. Of the major countries of origin, Syrian migrants are especially highly educated. In terms of our theoretical framework, this can be expected to reflect higher risks of staying in Syria due to high conflict intensity, and higher recognition rates of Syrian asylum applicants.

Fig. 3 illustrates the reasons for leaving by nationality. We find that more than 90 percent of respondents from Eritrea, Iraq, Somalia, Sudan, and Syria report leaving their country due to conflict or persecution. At the other end of the scale, a vast majority of respondents from Morocco and Algeria cite economic conditions as the main reason for leaving their home country. Additional analysis suggests that there are no noteworthy gender- and age-specific differences in reported reasons.

Finally, respondents were asked about the total cost of the journey and the total time spent traveling. In the Eastern (Central)

Mediterranean Route, 7 (15) percent reported total cost below 1000 USD, 74 (44) percent 1000–5000 USD, and 16 (12) percent above 5000 USD, with 3 (19) percent not being able or willing to give a cost estimate. In Turkey, costs were considerably lower, with 52 percent reporting that they had spent less than 1000 USD, 18 percent 1000–5000 USD, and only 4 percent more than 5000 USD. When it comes to the duration of travel, we find that 59 (65) percent of survey respondents in the Eastern (Central) Mediterranean Route spent less than 30 days in transit. 18 (20) percent of the respondents spent between 30 and 59 days, and 10 (13) percent of the respondents spent between 60 and 119 days.

6. Empirical approach

To assess the self-selection of migrants, we estimate a series of multivariate regression models relating to how refugees and irregular migrants in the FMS differ from the overall population in Gallup World Polls (GWP). We restrict our attention to those aged 25 to 64 to focus on

Table 2
Descriptive characteristics from Gallup World Polls, 2009–2014.

	(1) Full sample (all ages)	(2) Full sample (all ages) males	(3) Full sample (all ages) females
Age	35.22 (14.68)	36.14 (15.23)	34.23 (14.00)
Male	0.51 (0.50)	–	–
Secondary education	0.42 (0.49)	0.45 (0.50)	0.40 (0.50)
Tertiary education	0.08 (0.27)	0.10 (0.30)	0.06 (0.24)
Employed	0.48 (0.50)	0.65 (0.48)	0.29 (0.45)
Married	0.61 (0.49)	0.60 (0.48)	0.63 (0.48)
Divorced	0.02 (0.15)	0.02 (0.13)	0.03 (0.18)
Widowed	0.04 (0.21)	0.02 (0.14)	0.08 (0.27)
N	129,431	67,167	62,264
	Restricted sample (ages 25–64)	Restricted sample (ages 25–64) males	Restricted sample (ages 25–64) females
Age	39.31 (10.90)	39.95 (11.13)	38.63 (10.61)
Male	0.52 (0.50)	–	–
Secondary education	0.37 (0.48)	0.40 (0.49)	0.35 (0.48)
Tertiary education	0.10 (0.29)	0.12 (0.32)	0.07 (0.26)
Employed	0.55 (0.50)	0.75 (0.43)	0.33 (0.47)
Married	0.76 (0.42)	0.77 (0.42)	0.76 (0.43)
Divorced	0.03 (0.17)	0.02 (0.15)	0.04 (0.20)
Widowed	0.04 (0.21)	0.02 (0.15)	0.08 (0.71)
N	89,484	46,493	42,991

Notes: Means (standard deviations). This table presents summary statistics for origin countries included in the analysis. Secondary and tertiary education refer to the highest completed education. Source: Gallup World Polls, 2009–2014.

individuals most likely to have completed their education, and not yet retired. Our main variables of interest are age, gender, and level of education. In some of the analyses, we also use predicted incomes to study how these are related to self-selection into emigration. We proceed by estimating linear probability models for our outcomes for ease of interpretation, though logistic regression models returned similar patterns. We also estimate most models separately by gender.

Our models of self-selection take the form:

$$\text{Migrant}_{ic} = \alpha + \beta_1 X_i + \beta_2 O_c + \varepsilon_i \quad (3)$$

where Migrant_{ic} takes a value of 1 if individual i from country c is in the FMS sample and 0 otherwise. X_i is a vector of demographic variables that include: age group dummies (25–34, 35–44, 45–54 in our main analysis); education dummies (tertiary education, secondary education); dummy variable indicating the labor market status before migration (employed); a dummy variable indicating the marital status (married, divorced, widowed). To account for other unobservable characteristics, we include a full set of origin country (O_c) dummies, which control for all time-invariant variation caused by factors that vary cross-nationally. We cluster bootstrapped standard errors at the origin country level.

We use data for the origin population from GWP between 2009 and 2014. However, the results are robust to using data between 2009 and 2011 (that is, pre-Syrian conflict period). Note that in equation (3) the relevant reference category is composed of individuals who are in the GWP.

We also use GWP to investigate the self-selection of refugees and

irregular migrants with respect to predicted income. Specifically, we first estimate Mincer regressions for each origin country controlling for the level of education, employment status, gender, and age. We then use these estimates to predict individual income for those surveyed in the FMS to gain further insights into how migrants from different country groups are self-selected with respect to their earnings potential.

7. Self-selection of refugees and irregular migrants into Europe

This section presents three sets of results. We first show self-selection of refugees and irregular migrants with respect to education and other demographic characteristics. Second, we analyze the self-selection of refugees and irregular migrants according to their predicted earnings. Third, we conduct several additional analyses to investigate the robustness to omitted variables bias and non-standard sampling framework.

7.1. Self-selection of refugees and irregular migrants with respect to education

In this subsection, we test our theoretical predictions on how the distribution of education among refugees and irregular migrants compares with that of non-migrants. We conduct this analysis in three different ways of distinguishing refugees and irregular migrants. First, we use individuals' responses to the "main reason to migrate" question in FMS. Second, we classify countries based on their country-level conflict intensity following the classifications provided by the UCDP.

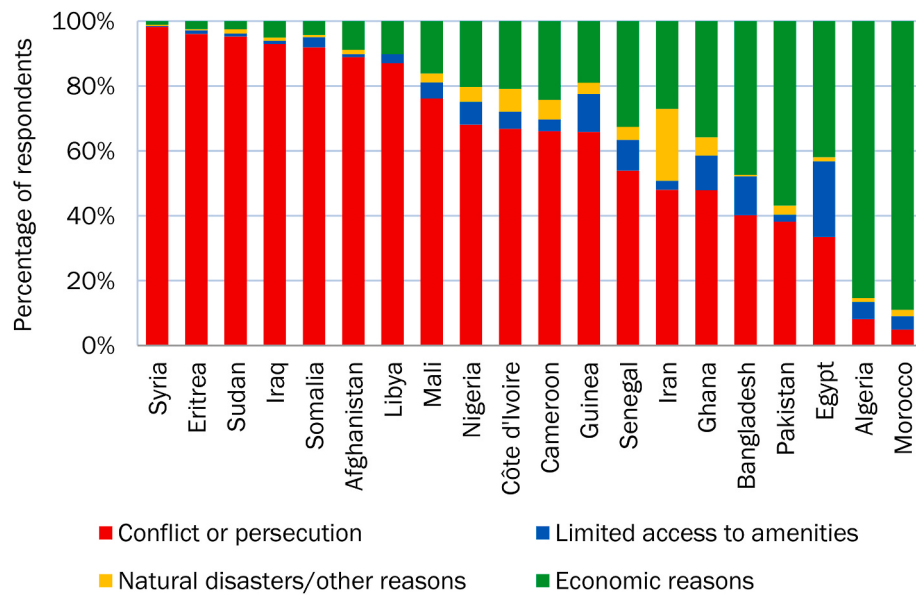


Fig. 3. Main reason to emigrate by origin country, age 14 and older.
Source: Flow Monitoring Surveys, 2015 and 2016 (waves 1–3) and authors' calculations.

Table 3
Self-selection of refugees and irregular migrants, adults aged 25–64.

Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)
→	Migrant due to any reason	Refugee	Irregular migrant	Refugee	Irregular migrant	Refugee	Irregular migrant
Sample	Both genders	Both genders	Both genders	Males	Males	Females	Females
→							
Secondary education	0.022*** (0.002)	0.029*** ^A (0.002)	−0.007*** (0.001)	0.020*** ^A (0.003)	−0.013*** (0.002)	0.026*** ^A (0.002)	−0.001 (0.001)
Tertiary education	0.037*** (0.004)	0.028*** ^A (0.003)	0.011*** (0.002)	0.035*** ^A (0.005)	0.004 (0.003)	0.035*** (0.004)	0.025*** (0.003)
Employed	0.001 (0.002)	0.000 (0.002)	0.002* (0.001)	0.018*** (0.003)	−0.004** (0.002)	−0.016*** (0.002)	0.004*** (0.001)
Male	0.057*** (0.002)	0.037*** (0.002)	0.016*** (0.001)	—	—	—	—
Age 25–34	0.079*** (0.002)	0.053*** (0.002)	0.026*** (0.001)	0.065*** (0.003)	0.035*** (0.002)	0.024*** (0.003)	0.012*** (0.001)
Age 35–44	0.035*** (0.002)	0.025*** (0.002)	0.011*** (0.001)	0.025*** (0.003)	0.018*** (0.002)	0.013*** (0.002)	0.002** (0.001)
Age 45–54	0.012*** (0.002)	0.008*** (0.002)	0.003*** (0.001)	0.008*** (0.003)	0.006*** (0.001)	0.003 (0.002)	−0.000 (0.001)
Married	−0.035*** (0.003)	−0.010*** (0.003)	−0.014*** (0.002)	−0.028*** (0.004)	−0.019*** (0.003)	0.024*** (0.002)	0.001 (0.002)
Divorced	−0.036*** (0.006)	−0.032*** (0.004)	0.005 (0.004)	−0.069*** (0.008)	0.002 (0.007)	0.009* (0.005)	0.015*** (0.005)
Widowed	−0.013*** (0.005)	−0.000 (0.004)	−0.006** (0.002)	−0.051*** (0.010)	−0.008 (0.005)	0.029*** (0.004)	0.004 (0.002)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
r ²	0.081	0.077	0.024	0.082	0.032	0.071	0.020
N	62,488	59,559	56,436	30,910	28,412	28,649	28,024

Notes: Robust standard errors in parentheses. Outcome variable, *migrant*, is equal to 1 for respondents in the Flow Monitoring Surveys and 0 for participants in Gallup World Polls. Reference categories are as follows: less than secondary education, unemployed or out of labor force, female, age 55+, and single. Source: Flow Monitoring Surveys, 2015 and 2016 (waves 1–3). Gallup World Polls, 2009–2014. ^A Indicates a statistically significant difference in each pair of estimates on the effects of education between refugees and irregular migrants at $p < .05$.

*** $p < .01$, ** $p < .05$, * $p < .1$.

Third, we match migrants' origin provinces to the sub-regions included in the GWP and define each sub-region's conflict intensity using geo-coded conflict data.

Using individuals' main reason for migration, Table 3 presents evidence for males and females together (Columns 1–3), for males only (Columns 4 and 5), and for females only (Columns 6 and 7), restricting the analysis of migrants to the first three waves. Outcome variables are as follows: *migrant due to any reason* is the outcome variable in Column 1;

refugee (which refers to migrants who left their country of origin due to conflict or persecution) in Columns 2, 4, and 6; *irregular migrant* (which refers to migrants who left their country of origin due to economic or other reasons, apart from conflict or persecution) in Columns 3, 5, and 7. In each analysis, the control group of non-migrants consists of GWP respondents. In each table, we present two-sided tests on whether the coefficients on secondary and tertiary education dummies are statistically significantly different across regressions for refugees and irregular

migrants. Results concerning statistical significance would remain unchanged if using one-sided tests for the hypothesis that the coefficients on secondary and tertiary education dummies are higher for refugees than for irregular migrants.

In column 1 of Table 3, we find that educated people are significantly more likely to migrate. Columns 2 and 3 show that the pattern is mainly driven by migrants escaping conflict or persecution.²⁵ Irregular migrants are also somewhat more likely to have completed tertiary education than non-migrants but to a lesser degree.²⁶ The difference in self-selection is statistically significant for both secondary and tertiary education.²⁷ Stronger positive self-selection of refugees is in line with our theoretical predictions suggesting that if the risk of being a victim of conflict or persecution increases, the probability of emigration becomes eventually increasing in human capital even if returns to human capital would be higher in the country of origin in the absence of conflict or persecution. Turning to other covariates, we find that the probability of emigration is higher for men, younger people, and singles. Young people being most likely to emigrate is intuitive as they have the longest time horizon to reap returns to migration investment (Sjaastad, 1962).

Columns 4–7 present corresponding analyses separately for men and women. Refugees are positively self-selected in terms of their education among both genders. Men who left due to other reasons (mostly economic) are negatively self-selected in terms of their education. This is intuitive and in line with Borjas (1987), as African and Asian countries have much wider income differences than European countries. The difference in self-selection between male refugees and male irregular migrants is statistically significant for both secondary and tertiary education. The negative self-selection of male irregular migrants compared with non-migrants, instead, is driven by fewer migrants with secondary education.

Women who left due to other reasons (mostly economic) are, instead, positively self-selected in terms of their education. Our theoretical framework can explain this in two ways. First, gender discrimination is a much bigger problem in Africa and Asia than in Europe. Together with very low female labor force participation rates in most, but not all, origin countries this can lower expected returns to education for women below expected returns in Europe, even when the opposite holds for men.²⁸ An alternative explanation would be that a non-negligible subset of women could be escaping gender-based violence and repression even from countries that are conventionally considered “safe”. This would produce a self-selection pattern that is similar to refugees who escape conflict and persecution. Still, female refugees are more strongly positively self-selected than female irregular migrants, although the difference is statistically significant only with respect to secondary education.

A third potential explanation, but outside our framework, is that there may be systematic gender differences in how expectations are formed or in the preciseness of such expectations. Previous research by McKenzie et al. (2013) suggests that such differences could be important. When surveying Tongans who applied to migrate to New Zealand through an immigration visa lottery both before and after migration, they found that males significantly under-estimated their post-migration

earnings, while females had reasonably accurate expectations. Unfortunately, we are not able to test the role of expectations as the FMS did not collect such information.

The decision of whether or not to migrate is a complicated one. It is often difficult to distinguish between refugees and those migrating for other reasons but claiming to be refugees. Economic incentives can be expected to play an important role even for those leaving countries suffering from civil war. Table 4 sheds light on this issue and presents the results by the conflict intensity classification at the country level. The columns are structured as follows: both genders and major conflict countries (column 1); both genders and minor or no conflict countries (column 2); males and major conflict countries (column 3); males and minor or no conflict countries (column 4); females and major conflict countries (column 5); and females and minor or no conflict countries (column 6). We again restrict attention to individuals aged 25 to 64.

The results in column 1 and 2 suggest that the probability of emigration increases in tertiary education in both country groups, but decreases in secondary education in countries with a minor or no conflict. In line with our results in Table 3, we find that migrants escaping major conflicts tend to be highly educated relative to the national average in their country of origin. When it comes to gender differences, we find that men and women emigrating from major conflict countries and women emigrating from minor or no conflict countries are positively self-selected in terms of their education, while men emigrating from minor or no conflict countries are negatively self-selected. Among both men and women, migrants from major conflict countries are statistically significantly more positively self-selected with respect to their education than migrants from minor or no conflict countries.

7.1.1. Could omitted country characteristics bias our results?

Of course, the presence of some unobserved reasons could affect both the probability for a country to be in conflict and the skill composition of emigrants. For example, poor institutional quality could both increase the risk of conflict and generate high-educated emigration due to bad job market prospects. Such unobserved heterogeneity might lead us to find that migrants from major conflict countries are positively self-selected. Furthermore, one can always ask whether the Uppsala Conflict Database cut-off of 1000 or more battle-related deaths within a year is an appropriate measure for high-level country risks. To address these concerns, we make use of the fact that we observe the origin province information for a subset of migrants surveyed in the FMS. More specifically, we match each of the migrants' origin provinces to the sub-regions included in the GWP to investigate whether our results hold when we conduct our analysis at the sub-regional level.²⁹ Formally, we proceed as follows: (i) we scrape geo-coded conflict data from Uppsala Conflict Database for the 2009–2014 period; (ii) we assign the number of battle-related deaths for each year to the corresponding sub-regions in the GWP based on geographical coordinates; (iii) we then classify each GWP sub-region as having *more intensive conflict* or *less intensive conflict*. *More intensive conflict* refers to the sub-regions that experience equal to or more battle-related deaths than the country median (over the 2009–2014 period), and *less intensive conflict* refers to the sub-regions

²⁵ We also find similar results when we analyze migrants arriving from the Eastern Mediterranean and the Central Mediterranean routes separately.

²⁶ The results hold also if excluding Iran, which has by far the highest share of tertiary educated emigrants.

²⁷ Unfortunately, we lack data for analyzing other forms of irregular immigration, like visa overstays. Since there are significant hurdles to obtain even a temporary visa, visa overstayers could be expected to be positively self-selected in terms of their education.

²⁸ In the age group 25 or older, female labor force participation rate varies between 15 percent and 33 percent and male labor force participation rate varies between 75 percent and 92 percent in Afghanistan, Algeria, Bangladesh, Iran, Iraq, Morocco, Pakistan, Sudan, and Syria (ILO 2019), with the gender gap in labor force participation rates varying between 53 and 68 percentage points. The gender gap in labor force participation rates is lower in Cameroon, Côte d'Ivoire, Nigeria, and Senegal, varying between 12 and 29 percentage points.

²⁹ Gallup defines sub-regions based on official boundaries. To address potential concern that GWP might be unrepresentative for conflict-affected sub-regions, we calculated GWP respondent shares and population shares based on censuses in those sub-regions covered by GWP in 2010 for both major and minor conflict countries. Average correlation across all countries is 0.941. Correlation is not only remarkably high for major conflict countries (0.983 for Afghanistan, 0.966 for Iraq and 0.987 for Syria) but also across high conflict intensive and low conflict intensive sub-regions. The full list of sub-regional population shares is available upon request.

Table 4

Self-selection of migrants from major and minor/no conflict countries, adults aged 25–64.

Sample →	(1) Both genders & Major conflict countries	(2) Both genders & Minor/no conflict countries	(3) Males & Major conflict countries	(4) Males & Minor/no conflict countries	(5) Females & Major conflict countries	(6) Females & Minor/no conflict countries
Secondary education	0.045*** ^A (0.004)	−0.006** (0.002)	0.045*** ^A (0.005)	−0.010** (0.004)	0.046*** ^A (0.004)	0.001 (0.002)
Tertiary education	0.049*** ^A (0.005)	0.016*** (0.005)	0.053*** ^A (0.007)	0.008 (0.008)	0.057*** ^A (0.007)	0.028*** (0.006)
Employed	0.003 (0.003)	−0.001 (0.003)	0.029*** (0.005)	−0.011** (0.005)	−0.034*** (0.003)	0.015*** (0.003)
Male	0.066*** (0.003)	0.044*** (0.003)	—	—	—	—
Age 25–34	0.097*** (0.004)	0.061*** (0.003)	0.126*** (0.006)	0.095*** (0.005)	0.050*** (0.005)	0.017*** (0.002)
Age 35–44	0.052*** (0.004)	0.016*** (0.002)	0.068*** (0.005)	0.021*** (0.003)	0.027*** (0.005)	0.001 (0.001)
Age 45–54	0.021*** (0.003)	0.005*** (0.001)	0.028*** (0.005)	0.005** (0.002)	0.009* (0.005)	−0.001 (0.001)
Married	−0.031*** (0.005)	−0.043*** (0.004)	−0.067*** (0.007)	−0.059*** (0.006)	0.039*** (0.005)	0.005 (0.003)
Divorced	−0.066*** (0.008)	−0.003 (0.008)	−0.133*** (0.013)	−0.001 (0.018)	0.016 (0.010)	0.025*** (0.008)
Widowed	−0.015** (0.007)	−0.018*** (0.005)	−0.088*** (0.016)	−0.033** (0.013)	0.049*** (0.008)	0.007* (0.004)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
r ²	0.079	0.061	0.082	0.072	0.068	0.028
N	34,405	28,083	18,869	14,384	15,536	13,699

Notes: Robust standard errors in parentheses. Outcome variable, migrant, is equal to 1 for respondents in the Flow Monitoring Surveys and 0 for participants in Gallup World Polls. Countries are classified by the level of conflict following the definitions provided by Uppsala Conflict Data Program: major conflict category includes countries with 1000 or more battle-related deaths in any of the years over the sample period (Afghanistan, Iraq, Nigeria, Pakistan, Sudan, Syria); minor conflict category includes countries with 25–999 battle-related casualties in any of the years over the sample period (Algeria, Iran); no conflict category includes countries that did not experience a major conflict or minor conflict in any of the years over the sample period (Bangladesh, Cameroon, Côte d'Ivoire, Morocco, Senegal). Reference categories are as follows: less than secondary education, unemployed or out of labor force, female, age 55+, and single. Source: Flow Monitoring Surveys, 2015 and 2016 (waves 1–3). Gallup World Polls, 2009–2014. ^A Indicates a statistically significant difference in each pair of estimates on the effects of education between migrants from major conflict and minor or no conflict countries at $p < .05$.

*** $p < .01$, ** $p < .05$, * $p < .1$.

that experience less battle-related deaths than the country median (over the 2009–2014 period).³⁰ Importantly, we do not include countries that did not experience any conflict, given that it is practically impossible to classify their sub-regions based on the conflict intensity.

We present the results at the sub-regional level in Table 5. The samples across the columns are as follows: both genders and more intensive conflict sub-regions (column 1); both genders and less intensive conflict sub-regions (column 2); males only and more intensive conflict sub-regions (column 3); males only and less intensive conflict sub-regions (column 4); females only and more intensive conflict sub-regions (column 5); and females only and less intensive conflict sub-regions (column 6).

The results in Table 5 are in line with findings from Tables 3 and 4: (i) migrants escaping from more intensive conflict sub-regions tend to be highly educated relative to the average in their sub-region of origin; (ii) men from more intensive conflict sub-regions are positively self-selected, whereas men from less intensive conflict sub-regions are negatively self-selected; (iii) women are positively self-selected from both less and more intensive conflict sub-regions, but more so from more intensive conflict sub-regions.³¹ Differences in self-selection are

statistically significant for both secondary and tertiary education among both men and women, as well as men and women pooled together.

Ideally, we would also have liked to use over-time variation in conflict intensity in origin countries. Unfortunately, our data do not allow this as FMS did not ask migrants their exact departure date, but asked them to choose an interval on how long ago they had left (less than 30 days, between 30 and 59 days, between 60 and 119 days, 120 days or more). As a result, we are only able to measure conflict intensity in the sub-region in a calendar year but not able to measure conflict intensity at the exact time of departure.

7.2. Self-selection with respect to predicted income

As the FMS did not ask respondents about their income, we can only test self-selection with respect to predicted incomes. We use predicted incomes for both migrants and non-migrants to focus on both groups' income arising from observable characteristics. We only focus on single individuals as income is reported at the household level in GWP. Specifically, we obtain predicted individual income by estimating within-country Mincer regressions which control for the level of education, employment status, gender, and age separately for each country of origin. We then use the estimated coefficients to calculate predicted income for both migrants and non-migrants. These results must be taken with great caution as migrants are likely to differ from non-migrants in several respects not captured in our data, as shown by McKenzie et al. (2010) in the context of Tongan migrants to New Zealand and non-migrants. Nonetheless, they provide the best estimates that the FMS allow us to derive.

Borjas et al. (2019) showed that the Roy model has more precise predictions about the self-selection of migrants than previously realized. The same conditions that result in positive or negative selection in terms

³⁰ The yearly numbers of observations by sub-regions are stable in all high conflict and minor conflict countries. Country median value over the 2009–2014 period is greater than zero in all countries.

³¹ Since we observe the origin province information for a subset of migrants surveyed in the FMS, we also checked whether using more granular conflict intensity variation leads qualitatively different findings and, consistent with our main results, we found that migrants escaping from more conflict intensive provinces tend to be highly educated. These results are not presented here but available upon request.

Table 5

Self-selection of migrants at the sub-regional level, adults aged 25–64, major and minor conflict countries only.

Sample →	(1) Both genders & More intensive conflict	(2) Both genders & Less intensive conflict	(3) Males & More intensive conflict	(4) Males & Less intensive conflict	(5) Females & More intensive conflict	(6) Females & Less intensive conflict
Secondary education	0.055*** ^Δ (0.007)	0.001 (0.006)	0.045*** ^Δ (0.009)	−0.018** (0.008)	0.077*** ^Δ (0.010)	0.025*** (0.008)
Tertiary education	0.091*** ^Δ (0.010)	0.013 (0.008)	0.096*** ^Δ (0.013)	−0.006 (0.011)	0.123*** ^Δ (0.017)	0.056*** (0.012)
Employed	0.067*** (0.007)	0.081*** (0.006)	0.167*** (0.010)	0.112*** (0.008)	−0.047*** (0.011)	0.040*** (0.008)
Male	0.069*** (0.007)	0.056*** (0.005)	—	—	—	—
Age 25–34	0.197*** (0.012)	0.126*** (0.008)	0.183*** (0.015)	0.127*** (0.011)	0.176*** (0.017)	0.105*** (0.011)
Age 35–44	0.112*** (0.012)	0.099*** (0.008)	0.085*** (0.015)	0.110*** (0.011)	0.121*** (0.017)	0.076*** (0.011)
Age 45–54	0.064*** (0.012)	0.037*** (0.008)	0.032** (0.016)	0.038*** (0.011)	0.081*** (0.018)	0.030*** (0.011)
Married	0.033*** (0.008)	−0.032*** (0.007)	−0.047*** (0.011)	−0.089*** (0.010)	0.177*** (0.012)	0.066*** (0.009)
Divorced	−0.050** (0.020)	0.067*** (0.018)	−0.232*** (0.032)	−0.101*** (0.028)	0.151*** (0.026)	0.221*** (0.022)
Widowed	0.060*** (0.018)	0.043*** (0.016)	−0.234*** (0.036)	−0.159*** (0.029)	0.254*** (0.021)	0.173*** (0.018)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
r ²	0.284	0.137	0.272	0.132	0.330	0.147
N	20,450	24,308	11,542	13,424	8908	10,884

Notes: Robust standard errors in parentheses. Sample only includes major and minor conflict countries: Afghanistan, Algeria, Iran, Iraq, Nigeria, Pakistan, Sudan, and Syria. Outcome variable, migrant, is equal to 1 for respondents in the Flow Monitoring Surveys and 0 for participants in Gallup World Polls. Sub-regions in Gallup World Polls classified as *more intensive conflict* or *less intensive conflict* based on the number of battle-related deaths data obtained from the Uppsala Conflict Data Program. *More intensive conflict* refers to the sub-regions that experience equal to or more battle-related deaths than the country median (over 2009–2014 period) and *less intensive conflict* refers to the sub-regions that experience less battle-related deaths than the country median (over 2009–2014 period). Reference categories are as follows: less than secondary education, unemployed or out of labor force, female, age 55+, and single. Source: Flow Monitoring Surveys, 2015 and 2016 (waves 1–3) and Gallup World Polls, 2009–2014. ^Δ Indicates a statistically significant difference in each pair of estimates on the effects of education between migrants from more intensive conflict sub-regions and migrants from less intensive conflict sub-regions at $p < .05$.

*** $p < .01$, ** $p < .05$, * $p < .1$.

of expected earnings also imply a stochastic dominance relationship between the earnings distributions of migrants and non-migrants.

Online Appendix Fig. A.3 compares cumulative distribution functions (CDFs) of predicted income (in the country of origin) of refugees who migrate due to conflict or persecution with non-migrants' CDF, and Fig. A.4 makes a corresponding comparison between irregular migrants whose main reason to emigrate was not conflict or persecution and non-migrants. These figures suggest that male and female refugees and female irregular migrants are positively self-selected with respect to their predicted income. Male irregular migrants, instead, do not differ much from non-migrants. Figs. A.5 and A.6 in the Online Appendix show that results are similar if analyzing self-selection based on country of origin: male and female migrants from major conflict countries and female migrants from minor or no conflict countries are positively self-selected in terms of their predicted income. Migrant and non-migrant men from countries with minor or no conflict do not differ in terms of their income distribution.

7.3. Could pseudo-cross-sectional sampling framework drive our results?

The sampling framework in our regression analysis is non-standard, in that the treatment and the control samples originate from different surveys. To address the concern that this could be driving our results, we also analyzed how self-selection of refugees differs from self-selection of irregular migrants when using only FMS data and including country fixed effects. In Table 6, we include all migrants in Flow Monitoring Survey waves 1–3. Our outcome variable is *migrant due to conflict or persecution*, which is equal to 1 for respondents who cite conflict or persecution as the main reason to migrate and 0 for respondents who

cite other reasons (economic reasons, limited access to amenities, or natural disasters and other reasons). The columns are structured as follows: all respondents (column 1); males (column 2); females (column 3). We again restrict our attention to individuals aged 25 to 64. We find that both male and female refugees are significantly more likely to have secondary and tertiary education compared with migrants who cite other reasons for leaving their countries. Finding consistent results within migrant only sample alleviates concerns that our results could be driven by the pseudo-cross-sectional sampling framework (i.e. migrant and source population samples come from different surveys) we use in our main analysis.

7.4. Robustness checks and additional results

7.4.1. Formally testing for omitted variables bias

Although sub-regional analysis presented above allows us to address some of the endogeneity concerns, one still might be concerned that conflicts may be more intense in areas with natural or strategic resources, or may cause more deaths in areas with high population density. These factors are can also be potentially correlated with the labor market structure, and so there may still be omitted variable bias. For example, if the returns to human capital are different in cities compared to rural areas, and conflict deaths occur more frequently in one of these types of regions, then this will affect the observed selection in conflict regions.

To investigate this concern formally, we perform a rigorous robustness check following the method proposed by Oster (2019). In top panel of Online Appendix Table A.6, we first reprint the baseline estimates from Table 5 for our main outcomes in the top rows for comparison

Table 6

Self-selection of refugees and irregular migrants, adults aged 25–64, Flow Monitoring Surveys sample only.

Outcome: <i>migrant due to conflict or persecution</i>	(1)	(2)	(3)
Sample	All	Males	Females
→			
Secondary education	0.067*** (0.013)	0.054*** ^A (0.015)	0.121*** (0.026)
Tertiary education	0.080*** (0.017)	0.087*** ^A (0.019)	0.048* (0.023)
Employed	−0.027** (0.014)	−0.020 (0.003)	−0.006 (0.034)
Male	0.053*** (0.007)	–	–
Age 25–34	0.176*** (0.034)	0.197*** (0.012)	0.112*** (0.022)
Age 35–44	0.102*** (0.022)	0.092*** (0.025)	0.088* (0.044)
Age 45–54	0.054** (0.022)	0.071** (0.032)	0.034*** (0.009)
Married	0.019 (0.014)	0.015 (0.015)	0.050 (0.049)
Divorced	−0.009 (0.042)	0.026 (0.057)	0.013 (0.072)
Widowed	0.040 (0.043)	0.068 (0.069)	0.031 (0.070)
Country FE	Yes	Yes	Yes
r ²	0.359	0.350	0.429
N	6473	4553	1920

Robust standard errors in parentheses. Outcome variable, migrant due to conflict or persecution, is equal to 1 for respondents who cite conflict or persecution as the main reason to migrate and 0 for other respondents who cite other reasons (economic reasons, limited access to amenities and natural disasters and other reasons) in the Flow Monitoring Surveys. Each specification includes fixed effects for survey waves and dummy variables for age 25–34, age 35–44 and age 45–54. Reference categories are as follows: less than secondary education, unemployed or out of labor force, female, age 55+, and single. Source: Flow Monitoring Surveys, 2015 and 2016 (waves 1–3). ^A Indicates a statistically significant difference in each pair of estimates on the effects of education between males and females at $p < .05$.

*** $p < .01$, ** $p < .05$, * $p < .1$.

purposes. The second panel presents the estimation bounds where we define R_{max} upper bound as 1.3 times the R-squared in specifications that control for observables following Oster (2019).³² The bottom panel presents the Oster's delta, which indicates the degree of selection on unobservables relative to observables that would be needed to fully explain our results by omitted variable bias.

The results presented point to a very limited movement in coefficients. High delta values also indicate that the unobservables have less effect on our coefficient of interest than the observables. Given the sub-regional variation we exploit and controls we include in our models, it is highly unlikely that there are unobserved factors that are 4–8 times as important as all observables. Therefore, the estimates suggest that our results are unlikely to be driven by omitted-variable bias.

7.4.2. Robustness to focusing on individuals aged 14 and above

We focus on 25–64 age group when analyzing self-selection as this

³² Estimation bounds on the treatment effect range between the coefficient from the main specification and the coefficient estimated under the assumption that observables are as important as unobservables for the level of R_{max} . R_{max} specifies the maximum R-squared that can be achieved if all unobservables were included in the regression. Oster (2019) uses a sample of 65 RCT papers to estimate an upper bound of the R-squared such that 90 percent of the results would be robust to omitted variables bias. This estimation strategy yields an upper bound for the R-squared, R_{max} , that is 1.3 times the R-squared in specifications that control for observables.

group is more likely to have completed their education. However, one might be concerned whether our results are driven by this choice. To address this formally, we present results focusing on individuals aged 14 and above in Online Appendix Table A.7. Our results remain qualitatively similar.

7.4.3. Testing the self-selection by GINI coefficient

As an additional test, we analyzed separately self-selection of male irregular migrants from more and less inequalitarian countries, based on whether the Gini coefficient in those countries is below or above source country sample median (see Online Appendix Table A.8). We find that negative self-selection or male irregular migrants is driven by negative self-selection from more inequalitarian countries, which is in line with Borjas (1987) and our theory.

7.4.4. Robustness to using origin population data from the 2009–2011 period

In Online Appendix Table A.9, we show that our results remain qualitatively the same when we use the data on the origin population between 2009 and 2011 (i.e. pre-Syrian conflict period).

7.4.5. Sub-regional conflict intensity and shares of emigration due to conflict or persecution

We also provide descriptive statistics on the share of those who emigrated due to conflict or persecution among migrants from sub-regions with more intensive and less intensive conflict, relative to the country median. Online Appendix Table A.10 reports this separately for men and women from all major and minor conflict countries. With the exception of women from Algeria and Nigeria, the share emigrating due to conflict or persecution is always higher from regions with more intensive conflict than from regions with less intensive conflict. For most countries, the difference is substantial among both men and women, suggesting that the major difference in self-selection pattern goes together with major differences in main motivations to emigrate, even when controlling for country fixed effects. Note that the overall share having emigrated due to conflict or persecution is lower in Online Appendix Table A.10 than among respondents included in Table 3. The discrepancy arises as sub-regions are observed only for a subset of respondents, and somewhat less common for those who emigrated due to conflict or persecution.

7.4.6. Do our results hold also in other settings?

So far, our econometric analysis has always relied on Flow Monitoring Surveys that interviewed migrants in transit countries. In Online Appendix Table A.11, we complement our analysis by investigating the self-selection of refugees into Germany, which is the main destination country, using data collected in the IAB-BAMF-SOEP Survey of Refugees (2016). By construction, the target group in that survey are refugees so we are not able to analyze self-selection of irregular migrants due to the low number of observations. Therefore, we only focus on migrants from Afghanistan, Iraq, Pakistan, and Syria (i.e. major conflict countries). Consistent with our main results, we find strong positive self-selection with respect to education.

In Online Appendix Table A.12, we also investigate the self-selection of refugees and irregular migrants with respect to education and other demographic characteristics in Turkey, which not only serves as a transit country for migrants but also hosts the largest refugee population in the world. We restrict attention to individuals aged 25 to 64 and to origin countries with more than 100 respondents in this age group in Turkey. The results for men and women together are, if anything, even more pronounced than in Table 3: refugees are strongly positively self-selected with respect to their education, and irregular migrants strongly negatively self-selected. The results for men are similar to Table 3: male refugees are positively self-selected, and male irregular migrants negatively self-selected. We also find that female refugees are positively self-selected, in line with Table 3. Unlike in Europe, female

irregular migrants in Turkey are not positively self-selected. In terms of our theoretical framework, this can be explained by worse gender discrimination and lower female labor force participation in Turkey.

7.4.7. Could liquidity constraints alone explain our results?

An additional mechanism outside our theoretical framework is that better-educated individuals would find it easier to finance their trip, while liquidity constraints would prevent the least educated people from migrating. However, if liquidity constraints would explain different self-selection patterns among refugees and irregular migrants, we would expect these groups to differ in their total spending to finance their trip. Additional analysis suggests that the distribution of total costs is almost the same among refugees and irregular migrants along both Eastern Mediterranean Route and Central Mediterranean Route (see Online Appendix Table A.13). Therefore, we find it unlikely that liquidity constraints would explain differences in self-selection patterns between these groups.

To further investigate to what extent liquidity constraints would be able to explain the patterns we observe, we re-estimate Table 6 while directly controlling for the cost of journey between the origin country and the country of interview. To do so, we create dummy variables for: (i) less than 1000 USD; (ii) between 1000 and 2500 USD; (iii) between 2500 and 5000 USD and (iv) more than 5000 USD as reported in the FMS. The results are presented in Online Appendix Table A.14. We find that the cost of journey does not predict who are refugees and who are irregular migrants neither for all migrants together, nor for men or women separately. Predictive power of education remains unaffected.

Last, we separately analyze the self-selection of migrants from poorest major conflict countries (Afghanistan, Nigeria, and Sudan), according to individually stated main reason to emigrate. Our theory suggests that male irregular migrants are negatively self-selected, while if everyone would like to emigrate from major conflict countries and liquidity constraints were decisive in who can emigrate, then also male irregular migrants should be positively self-selected. Online Appendix Table A.15 shows that the results are in line with our theory: male irregular migrants are negatively self-selected also from poorest major conflict countries, while refugees and female irregular migrants are positively self-selected. Again, the differences in self-selection with respect to education are statistically significant. Our theory can explain all these patterns, while the competing hypothesis that self-selection patterns would be driven by liquidity constraints cannot explain why male irregular migrants from poorest major conflict countries are negatively self-selected.

7.4.8. Could the distance between the origin country and destination country explain our results?

If migrants fleeing war and persecution had covered longer distance, a potential concern would be that this could explain the positive selection in terms of education (as greater geographical distance between origin and destination country makes a difference to costs). To address this concern we directly control for travel costs and the great-circle distance between the two largest cities of each country pair in FMS in Online Appendix Table A.16 (that is, the distance between the migrants' origin country and intended destination country in a sub-sample of respondents who reported both). Overall, controlling for the observed cost of the trip as well as geographical distance does not change the main positive self-selection pattern we found. These findings suggest that liquidity constraints alone do not drive self-selection patterns in terms of education neither for all migrants together, nor for men or women separately. Even more, we find that refugees travel, on average, a shorter distance.

7.4.9. Are the results driven by the migrant network effect?

McKenzie and Rapoport (2010) and Guichard (2020) have shown the critical role of migration networks in explaining the self-selection of migrants in terms of education. Therefore, a potential threat to our main

findings arises if more educated migrants also have existing networks abroad. If that were the case, our results could potentially be explained by the migrant network effect.

To alleviate this concern, we directly control for the presence of a network in Online Appendix Table A.16 (in a sub-sample of respondents, in which we know migrants' origin country and intended destination country). We do this in two ways: (i) FMS asks the reasons behind the respondents' reported intended destination countries. We code options "family/relatives there" and "network of co-nationals already established there" as 1 if the migrants chose these options and zero otherwise; (ii) we control for (log) diaspora population in the intended destination country. Adding these additional "migrant network" controls does not change our main results in a meaningful way.

7.4.10. Does the difference in selection for irregular migrants vary by marital status?

We finally investigate whether the different selection of irregular migrants by gender depends on marital status. In particular, using individuals' main reason for migration (similar to Table 3) we present in Online Appendix Tables A.17 and A.18 evidence for single and married males, and for single and married females. In each table, we present two-sided tests on whether the coefficients on secondary and tertiary education dummies are statistically significantly different across regressions for married and single migrants of the studied gender.

The results reported in Online Appendix Table A.17 indicate that single male irregular migrants are strongly negatively self-selected in terms of their education. Married male irregular migrants, on the other hand, are somewhat more likely to have completed tertiary education than non-migrants. The difference in self-selection between single and married male irregular migrants is statistically significant for both secondary and tertiary education. Overall, the results suggest that negative male self-selection reported in Table 3 is mostly driven by single male irregular migrants.

When it comes to female irregular migrants, we again find that both single and married women are positively self-selected in terms of their education. These results are in line with our theoretical framework, which suggests that gender discrimination as well as gender-based violence and repression can lower expected returns to education for both single and married women.

Finally, in Online Appendix Table A.18, we investigate whether self-selection of refugees vary by marital status. We find that both male and female refugees are positively self-selected with respect to their education, irrespective of their marital status.

8. Conclusion

In this paper, we analyzed self-selection of refugees and irregular migrants. We found that refugees and female irregular migrants who arrived in Europe are positively self-selected in terms of their education and expected earnings, while male irregular migrants are negatively self-selected. This pattern holds both when using individually stated main reason to emigrate, and when comparing migrants from major conflict countries (most plausibly refugees) with migrants from minor or no conflict countries. To address the concern that both emigration and conflict could reflect unobserved country-specific factors, we also used geo-coded conflict data to classify sub-regions in each conflict country according to their conflict intensity. Our findings on different self-selection patterns hold also in this sub-regional analysis, alleviating concerns that the patterns would be driven by cross-country differences in colonial history, institutions, or the presence of oil and other natural resources. The differences in self-selection with respect to education are statistically significant in line with our theory.

To evaluate whether the patterns we found could reflect liquidity constraints, we analyzed self-selection from poorest major conflict countries. Male irregular migrants (those who stated that their main reason to emigrate was something else than conflict or persecution) are

negatively self-selected also from poorest major conflict countries, in line with our theory, but contrary to what one would expect if self-selection would reflect conflict intensity and liquidity constraints. Therefore, our theory can explain how self-selection differs between refugees and irregular migrants among both men and women, while the hypothesis that liquidity constraints drive differences in self-selection cannot explain why male irregular migrants even from poorest major conflict countries are negatively self-selected.

Positive self-selection of refugees is good news for receiving countries if they keep taking steps to ease their path to labor market integration. For example, successful integration of refugees can boost subsequent exports to their country of origin (Parsons and Vézina, 2018). At the same time, the positive self-selection of refugees is bad news for countries of origin as it makes rebuilding more difficult, once the conflict is over. To the extent that human capital is driving economic growth, such a loss may have severe long-term effects.

Our findings also imply that policies that support refugees' and migrants' integration into the labor market need to be tailored to their skills (World Bank, 2018). Migrants escaping major conflicts (such as the fighting in Syria) may well benefit from receiving early access to language courses and other basic training while waiting for decisions on their asylum applications. Battisti et al. (2019) show that the labor market outcomes of certain groups of refugees can be improved through a basic job search assistance program. Along the same lines, Brell et al. (2020) find that keeping the asylum process short, providing early support to address health issues, and facilitating refugees to join the labor market at the earliest possible stage are of key importance for refugees' integration. All of these aspects are important in terms of promoting the integration of migrants and helping a country to make society more inclusive for everyone.

Authors statement

We, as authors of the manuscript "Refugees' and Irregular Migrants' Self-selection into Europe", declare no conflict of interest. All authors contributed equally.

Data availability

The authors do not have permission to share data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jdeveco.2021.102681>.

References

- Abramitzky, R., Boustan, L., Eriksson, K., 2012. Europe's tired, poor, huddled masses: self-selection and economic outcomes in the age of mass migration. *Am. Econ. Rev.* 102 (5), 1832–1856.
- Ajzenman, N., Aksoy, C.G., Guriev, S., 2020. Exposure to Transit Migration, Public Attitudes and Entrepreneurship. CEPR Discussion Paper DP14605.
- Aksoy, C.G., Poutvaara, P., Schikora, F., 2020. First Time Around: Local Conditions and Multi-Dimensional Integration of Refugees. IZA DP No. 13914.
- Aksoy, C.G., Tumen, S., 2021. Local governance quality and the environmental cost of forced migration. *J. Dev. Econ.* 149, 102603.
- Barone, G., D'Ignazio, A., de Blasio, G., Naticchioni, P., 2016. Mr. Rossi, Mr. Hu and politics. The role of immigration in shaping natives' voting behavior. *J. Publ. Econ.* 136, 1–13.
- Battisti, M., Giesing, Y., Laurentsyeve, N., 2019. Can job search assistance improve the labour market integration of refugees? *Lab. Econ.* 61 article number 101745.
- Belot, M.V., Hatton, T.J., 2012. Immigrant selection in the OECD. *Scand. J. Econ.* 114 (4), 1105–1128.
- Besley, T., Persson, T., 2011. The logic of political violence. *Q. J. Econ.* 126 (3), 1411–1445.
- Birgier, D.P., Lundh, C., Haberfeld, Y., Elldér, E., 2018. Self-selection and host country context in the economic assimilation of political refugees in the United States, Sweden, and Israel. *Int. Migrat. Rev.* 52 (2), 524–558.
- Borjas, G.J., 1987. Self-selection and the earnings of immigrants. *Am. Econ. Rev.* 77 (4), 531–553.
- Borjas, G.J., Kauppinen, I., Poutvaara, P., 2019. Self-selection of emigrants: theory and evidence on stochastic dominance in observable and unobservable characteristics. *Econ. J.* 129 (617), 143–171.
- Brell, C., Dustmann, C., Preston, C., 2020. The labor market integration of refugee migrants in high-income countries. *J. Econ. Perspect.* 34 (1), 94–121.
- Brücker, H., Rother, N., Schupp, J., von Gostomski, C.B., Böhm, A., Fendel, T., Friedrich, M., Giesselmann, M., Kosyakova, Y., Kroh, M., Kühne, S., Liebau, E., Richter, D., Romiti, A., Schacht, D., Scheible, J.A., Schmelzer, P., Siegert, M., Sirries, S., Trübawetter, P., Vallizadeh, E., 2016. Forced migration, arrival in Germany, and first steps toward integration. In: Britzke, J., Schupp, J. (Eds.), *SOEP Wave Report 2016*, pp. 149–164.
- Chin, A., Cortes, K.E., 2015. The refugee/asylum seeker. In: Chiswick, B.R., Miller, P.W. (Eds.), *Handbook of the Economics of International Migration*, volume 1, pp. 585–658.
- Chiquiar, D., Hanson, G.H., 2005. International migration, self-selection, and the distribution of wages: evidence from Mexico and the United States. *J. Polit. Econ.* 113 (2), 239–281.
- Chiswick, B., 1999. Are immigrants favorably self-selected? *Am. Econ. Rev.* 89 (2), 181–185.
- Cohen, Y., 2007. Self-selection and earnings assimilation: immigrants from the former Soviet union in Israel and the United States. *Demography* 44 (3), 649–668.
- Dustmann, C., Fasani, F., Frattini, T., Minale, L., Schönberg, U., 2017. On the economics and politics of refugee migration. *Econ. Pol.* 32 (91), 497–550.
- Dustmann, C., Vasiljeva, K., Damm, A.P., 2019. Refugee migration and electoral outcomes. *Rev. Econ. Stud.* 86 (5), 2035–2091.
- Edo, A., Giesing, Y., Öztunc, J., Poutvaara, P., 2019. Immigration and electoral support for the far-left and the far-right. *Eur. Econ. Rev.* 115, 99–143.
- Ethier, W.J., 1986. Illegal immigration: the host-country problem. *Am. Econ. Rev.* 76 (1), 56–71.
- Eurostat, 2017. Data on Asylum Applicants. Available here: <https://ec.europa.eu/eurostat/news/themes-in-the-spotlight/asylum2016>.
- Fernandez-Huertas Moraga, J., 2011. New evidence on emigrant selection. *Rev. Econ. Stat.* 93 (1), 72–96.
- Friebel, G., Guriev, S., 2006. Smuggling humans: a theory of debt-financed migration. *J. Eur. Econ. Assoc.* 4 (6), 1085–1111.
- Gibney, M., Cornett, L., Wood, R., Haschke, P., Arnon, D., Pisanò, A., 2018. The Political Terror Scale 1976–2017. Available here: <http://www.politicalterror scale.org>. (Accessed 17 November 2019).
- Grogger, J., Hanson, G.H., 2011. Income maximization and the selection and sorting of international migrants. *J. Dev. Econ.* 95, 42–57.
- Guichard, L., 2020. Self-selection of asylum seekers: evidence from Germany. *Demography* 57 (3), 1089–1116.
- Halla, M., Wagner, A.F., Zweimüller, J., 2017. Immigration and voting for the far right. *J. Eur. Econ. Assoc.* 15 (6), 1341–1385.
- Hangartner, D., Dinas, E., Marbach, M., Matakos, K., Xefteris, D., 2019. Does exposure to the refugee crisis make natives more hostile? *Am. Polit. Sci. Rev.* 113 (2), 442–455.
- Hanson, G.H., Spillimbergo, A., 1999. Illegal immigration, border enforcement, and relative wages: evidence from apprehensions at the US-Mexico border. *Am. Econ. Rev.* 89 (5), 1337–1357.
- Hatton, T.J., 2016. Refugees, asylum seekers, and policy in OECD countries. *Am. Econ. Rev.* 106 (5), 441–445.
- Hatton, T.J., 2020. Asylum migration to the developed World: persecution, incentives, and policy. *J. Econ. Perspect.* 34 (1), 75–93.
- Hatton, T.J., Williamson, J.G., 1998. *The Age of Mass Migration: Causes and Economic Impact*. Oxford University Press.
- Hendricks, L., Schoellman, T., 2018. Human capital and development accounting: new evidence from wage gains at migration. *Q. J. Econ.* 133 (2), 665–700.
- ILO, 2019. Labour Market Statistics. Available here: <https://www.ilo.org/ilostat/>. (Accessed 17 November 2019).
- IOM, 2016. Mixed Migration Flows in the Mediterranean and beyond. Available here: http://migration.iom.int/docs/Analysis_Flow_Monitoring_and_Human_Trafficking_Surveys_in_the_Mediterranean_and_Beyond_8_DECEMBER_2016.pdf. (Accessed 17 November 2019).
- IOM, 2017. Mixed Migration Flows in the Eastern Mediterranean and beyond. Available here: http://migration.iom.int/docs/Analysis_Flow_Monitoring_Surveys_in_the_Mediterranean_and_Beyond_11_April_2017.pdf. (Accessed 17 November 2019).
- Lessem, R., 2017. Mexico-US immigration: effects of wages and border enforcement. *Rev. Econ. Stud.* 85 (4), 2353–2388.
- McKenzie, D., Gibson, J., Stillman, S., 2010. How important is selection? Experimental vs. non-experimental measures of the income gains from migration. *J. Eur. Econ. Assoc.* 8 (4), 913–945.
- McKenzie, D., Gibson, J., Stillman, S., 2013. A land of milk and honey with streets paved with gold: do emigrants have over-optimistic expectations about incomes abroad? *J. Dev. Econ.* 102, 116–127.
- McKenzie, D., Rapoport, H., 2010. Self-selection patterns in Mexico-US migration: the role of migration networks. *Rev. Econ. Stat.* 92 (4), 811–821.
- Orrenius, P.M., Zavodny, M., 2005. Self-selection among undocumented immigrants from Mexico. *J. Dev. Econ.* 78 (1), 215–240.
- Oster, E., 2019. Unobservable selection and coefficient stability: theory and evidence. *J. Bus. Econ. Stat.* 37 (2), 187–204.
- Otto, A.H., Steinhart, M.F., 2014. Immigration and election outcomes: evidence from city districts in Hamburg. *Reg. Sci. Urban Econ.* 45, 67–79.
- Parey, M., Ruhose, J., Waldinger, F., Netz, N., 2017. The selection of high-skilled emigrants. *Rev. Econ. Stat.* 99 (5), 776–792.
- Parsons, C., Vézina, P.L., 2018. Migrant networks and trade: the Vietnamese boat people as a natural experiment. *Econ. J.* 128 (612), F210–F234.

- Pettersson, T., Eck, K., 2018. Organized violence, 1989–2017. *J. Peace Res.* 55 (4), 535–547.
- Pew Research Center, 2019. 5 Facts about Illegal Immigration in the U.S. Available here: <https://www.pewresearch.org/fact-tank/2019/06/12/5-facts-about-illegal-immigration-in-the-u-s/>. (Accessed 17 November 2019).
- Roy, A.D., 1951. Some thoughts on the distribution of earnings. *Oxf. Econ. Pap* 3 (2), 135–146.
- Ruyssen, I., Salomone, S., 2018. Female migration: a way out of discrimination? *J. Dev. Econ.* 130, 224–241.
- Sjaastad, L.A., 1962. The costs and returns of human migration. *J. Polit. Econ.* 70 (5, Part 2), 80–93.
- Spitzer, Y., Zimran, A., 2018. Migrant self-selection: anthropometric evidence from the mass migration of Italians to the United States, 1907–1925. *J. Dev. Econ.* 134, 226–247.
- Steinmayr, A. (forthcoming). Contact versus Exposure: Refugee Presence and Voting for the Far-Right. *Review of Economics and Statistics*.
- UN, 2017. Gender Inequality Index. Available here: <http://hdr.undp.org/en/composite/GII>. (Accessed 17 November 2019).
- UNHCR, 1967. Convention and Protocol Relating to the Status of Refugees. Available here: <http://www.unhcr.org/3b66c2aa10>. (Accessed 17 November 2019).
- UNHCR, 2008. Challenges of Irregular Migration: Addressing Mixed Migration Flows. Available here: www.unhcr.org/49e479c911.pdf. (Accessed 17 November 2019).
- UNHCR, 2017. Global Trends: Forced Displacement in 2016. <https://www.unhcr.org/5943e8a34.pdf>. (Accessed 17 November 2019). Available here: <https://data2.unhcr.org/en/situations/mediterranean>. (Accessed 17 November 2019).
- UNHCR, 2018. Sea and Land Arrivals Monthly. Available here: <https://data2.unhcr.org/en/situations/mediterranean>. (Accessed 17 November 2019).
- UNHCR, 2019. Global Trends: Forced Displacement in 2018. Available here: <https://www.unhcr.org/5d08d7ee7.pdf>. (Accessed 17 November 2019).
- World Bank, 2018. Asylum Seekers in the European Union: Building Evidence to Inform Policy Making. Washington, D.C.