

Schmid, Lena; Renner, Laura

Conference Paper

The Decision to Flee: Analyzing Gender-Specific Determinants of International Refugee Migration

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2020: Gender Economics

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Schmid, Lena; Renner, Laura (2020) : The Decision to Flee: Analyzing Gender-Specific Determinants of International Refugee Migration, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2020: Gender Economics, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/224596>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The Decision to Flee: Analyzing Gender-Specific Determinants of International Refugee Migration*

Laura Renner[†] Lena Schmid[‡]

PRELIMINARY VERSION.

Please do not cite without permission.

February 2020

Abstract

Using a gravity approach, we explore determinants and dynamics of refugee migration using a gendered cross-national dyadic dataset on refugee movements for the years 2000-2015. Along three dimensions (push, pull and cost factors), we analyze whether there are heterogeneous effects of flight determinants for men and women. Our results suggest that within the push dimension there is only little difference in gender-specific responses. Most prominently, women react relatively stronger to the existence of sexual violence as a conflict strategy. When it comes to dyadic factors, we find that distance is a decisive factor for men and women. The most gender-sensitive dimension appears to be the pull factors: Neighboring countries attract more women in comparison to men as well as to non-neighbors. For non-neighbors, female flows are more sensitive to political stability, women's rights and the economic situation of the destination

JEL-Code: F22, O15, J16

Keywords: Refugees, International Migration, Distance, Conflict, Gender

*An earlier version of this paper entitled 'Refugee Migration: Determinants and Dynamics' was presented at the Frias Conference The Economics of Migration: Perspectives on Family, Gender, and Fertility, the Jan Tinbergen European Peace Science Conference in The Hague 2019, the Insights in Political Economy Workshop in Bayreuth 2019, the WomanStats Project Summer meeting 2019 and the Annual Conference of the European Economic Association in Manchester. The authors thank Tim Krieger, Mario Larch, Daniel Meierrieks and Valentin Klotzbücher as well as seminar participants at the RWI Essen for most helpful comments and suggestions.

[†]Department of Economics, University of Freiburg, Wilhelmstr. 1b, 79085 Freiburg i. Br., Germany; E-mail: laura.renner@vwl.uni-freiburg.de; Phone: +49 761 203 67653

[‡]Department of Economics, University of Freiburg, Wilhelmstr. 1b, 79085 Freiburg i. Br., Germany; E-mail: lena.schmid@vwl.uni-freiburg.de; Phone: +49 761 203 67652.

1 Introduction

20.4 million – the number of refugees counted by the United Nations High Commission for Refugees (UNHCR) in 2018 is the highest number ever reported (UNHCR, 2019). For the seventh year in a row numbers are rising, intensifying problems of international coordination and humanitarian aid. One third of the refugee population is located in Least Developed Countries which themselves face severe development problems and usually do not have the capacity needed to adequately support these people seeking refuge. Another 18% of all refugees are based in Turkey which thus hosts more displaced people than all European countries together. A further case to highlight is Lebanon which receives most refugees relative to its national population. For every 1,000 Lebanese people the country hosts 156 refugees.

Concerning the origin of refugees, only few countries are responsible for large parts of the refugee population. The UNHCR (2019) reports that two thirds of all refugees come from only five different countries. As in the past years, the largest share of refugees are still from Syria with most of them fleeing to Turkey or other countries in the area. Other source countries with large outflows of refugees are Afghanistan, South Sudan, Myanmar and Somalia.

Why is it that we observe exactly these movements? We know that they mostly escape from violence and persecution, but how do refugees actually decide whether and when they flee and how do they choose their country of destination? And further, are there heterogeneous effects for different groups of the refugee population? Although all people in a certain country experience conflict, it can be observed that some people do not leave the country as well as timing and destination choice varies substantially. In this exploratory study we examine whether there are gender differences in flight patterns, specifically we investigate if and how females and males react differently to determinants of refugee migration. Using a novel dataset with the demographic composition of dyadic refugee stocks from the UNHCR for the years 2000-2015, we are the first to study gender-specific aspects of refugee movements.

The relevance of the question is best explained when looking at the gendered composition of refugee flows in the two regions where most refugees live: Africa and Asia. For the period we observe, the mean share of females in total bilateral refugee flows is always below 0.5—that is women account for less than half of the refugee flows, as shown in Figure 1.

Figure 1 about here

When studying refugee migration one has to be aware that it is a specific form of migration (for an extended discussion of the terminology see section 2) where people have a rather low agency over their decision to leave their home country. While voluntary migration

(mostly labor migration) has been studied extensively, asking questions such as who leaves his or her home country and where do people go, this has rarely been done for refugee migration. Refugees predominantly flee from violence, conflict and persecution which is, roughly speaking, the criteria to be granted the status of a refugee under the United Nations Geneva Convention¹. On first sight, this seems to be a strong contrast to voluntary migration as the primary determinant is not maximizing life-time utility by improving (economic) opportunities but to secure survival. For this reason, flight migration is often classified as involuntary or forced migration. However, even in times of conflict and war, rarely do entire populations flee. This shows that even forced migrants—although within a much more limited scope than voluntary migrants—still have a choice to make about when and where to go. Accordingly, it can be observed that flight patterns are manifold (World Bank, 2017, pp.43-49). We argue that it is reasonable to ask where these differences come from as it seems that fleeing is not equally likely for all parts of the population. In this study we analyze the gendered composition of dyadic refugee flows to shed a light on the decision making of male and female refugees.

Thus, the underlying question is: Do women and men decide whether and where to flee in a similar fashion? Are, for example, costs arising from distance higher for women than for men? Are men and women attracted by different characteristics of destination countries? And is one of the genders, in general, more prone to leave the home country and under which circumstances?

We contribute to different strands of literature by transferring and adapting the standard model for migration to flight migration. The first contribution we can make to is the field of determinants of forced migration (Schmeidl, 1997; Adhikari, 2012; Moore and Shellman, 2004, 2006, 2007; Echevarria and Gardeazabal, 2016; Davenport et al., 2003; Iqbal, 2007; Neumayer, 2004, 2005; Rügger and Bohnet, 2018). While these studies examine what drives refugees out of their country and what attracts them to a certain destination, the majority (except for Rügger and Bohnet (2018) who study refugees with respect to their ethnicity) treats refugees as a homogeneous group and derives implications for the whole refugee population. With our data on the demographic composition we are able to disentangle potential heterogeneous effects of these determinants for males and females. The second body of literature this study can contribute to is the one analyzing gender-specific aspects of voluntary migration (Pedraza, 1991; Baudassé and Bazillier, 2014; Beine and Salomone, 2013; Docquier et al., 2009, 2012; Dumont et al., 2007; Nejad and Young, 2014; Ruysen and Salomone, 2018). Here, our paper can broaden the perspective and show similarities and differences for the different types of migrants. Section 3 provides a more detailed review of the relevant literature.

Our approach may be summarized by looking at two examples. We distinguish between

¹For a more detailed description of the legislation, as well as recent developments, see e.g. Dustmann et al. (2017, pp.501-508)

different dimensions of the decision to flee. First, we look at reasons to flee emerging from the situation in the home country. For example, we can see in Figure 2 showing refugee outflows by gender in Myanmar over time that outflows are mostly (though not in all years) dominated by men. Second, we look at the decision where to go, often described as sorting. Plotting yearly inflows by gender, Figure 3 shows that inflows into Egypt are again mostly male driven, while the difference between male and female inflows differs across years.

Figure 2 about here

Figure 3 about here

The importance of understanding why people leave and where they go including heterogeneous effects by gender is not only interesting in itself but of great interest for further analysis as well as policy implications. Refugee movements can heavily impact their country of origin and their host country at the same time. Such in- or outflow of people may have various implications for origin and host countries, such as brain drain (Docquier and Rapoport, 2012), increasing conflict in the country of asylum (Salehyan and Gleditsch, 2006; Gineste and Savun, 2019; Rüegger, 2019) or economic consequences in destinations (Ruiz and Vargas-Silva, 2015; Tumen, 2016; Borjas and Monras, 2017). These effects may vary depending on the demographic composition of the refugee population. Emigration of high-skilled females, for example, is found to have a negative impact on education and health in the origin country (Dumont et al., 2007). Furthermore, it is important for home and destination countries as skewed sex ratios might have consequences for labor markets, fertility but also potentially for violence and crime (for example, Hudson and den Boer (2004) discuss destabilizing effects of sex ratios with a male surplus in Asia). Lastly, knowing more about the demographic composition of refugee flows and the related, sex-specific, determinants is important because it sheds light on the possible consequences for destinations interested in integration (Dustmann et al., 2017).

The remainder of the paper is organized as follows. Section 2 explains the migration and refugee terminology and Section 3 summarizes the current related literature. Section 4 introduces standard migration theory and elaborates how it can be applied to the particular case of refugee migration. Section 5 introduces the data on refugee flows including demographic information, explains the sample used and provides information on the variables identified as push and pull factors for refugee migration. The estimation strategy is described in Section 6. Section 7 describes and discusses our preliminary results.

2 Who are we talking about

Who is a refugee? Who is a migrant? Can we apply migration theory to refugees and where are potential differences and similarities? To answer these questions, we first need

to know who we are talking about. According to the 1951 Geneva Convention refugees are people outside their home country “owing to well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion” (UNHCR 1951) . This is the definition of people who are legally protected by being granted asylum by countries that have signed the convention.²

Discussion prevails whether the legally protected refugees stand in contrast to or are subgroup of migrants. Several international organizations define a migrant simply as a person who moves away from his or her place of usual residence (IOM, 2019). Thus, migrant is an umbrella term for a group where refugees are part of. Carling (2017) refers to this as the ‘inclusivist approach’. Opposed to this perspective is the ‘residualist view’ which is prominently represented by the UNHCR. Following this approach, a clear distinction between refugees and migrants is necessary. It emphasizes the level of agency people have over their decision to leave their home, either being forced (refugees) or deciding voluntarily (migrants). This definition has also been acknowledged by the UN General Assembly in the New York Declaration for Refugees and Migrants.³

Unfortunately, there is no universal treatment of the terms migrants and refugees. In this paper we take up the inclusivist view and treat refugees as a subgroup of all migrants. According to our understanding migrants are people who reside in a country different from their country of birth. We further understand that the group of migrants includes forced migrants as well as voluntary migrants and the continuum that lies in between those. We see forced and voluntary migrants as the polar cases of a spectrum of agency over migration decision (see e.g. Richmond, 1993). Along this continuum, migrants have increasing room for decision about their migration process, however, even forced migrants have some - even if very little - degree of freedom in their decision when and where to flee.

3 Literature review

The first contribution we make is to the growing body of empirical studies that study determinants of forced migration. The evidence widely agrees that violent conflict in the home country is a significant predictor of positive refugee flows (see e.g. Schmeidl (1997); Moore and Shellman (2004); Iqbal (2007); Echevarria and Gardeazabal (2016)) while the costs of moving, i.e. distance, social networks or shared borders between the country of origin and the country of asylum, are important factors for the decision where to go (see e.g. Rügger and Bohnet (2018); Echevarria and Gardeazabal (2016); Moore and Shellman (2007); Iqbal (2007); Adhikari (2012); Neumayer (2004)).

²For a more detailed description of the legislation, as well as recent developments, see e.g. Dustmann et al. (2017, pp.501-508).

³For a thorough discussion of the debate about migrants and refugees see Carling (2015) and Carling (2017).

Among the first who studied drivers of refugee migration in an empirical cross-country setting was Schmeidl (1997). Using emigration data from the UNHCR she found that measures of political violence are better predictors for refugee numbers than institutional measures like human rights, for instance. The effect of violence is amplified if foreign military is involved in the conflict which shows the effect that, for example, Western states can indirectly exert on refugee flows. Further, geography-related facilitators or obstacles to flight like the number of bordering countries are not found to have a significant impact on refugee movements. One potential reason might be that the study does not measure dyadic refugee flows but the stock of refugees globally from a certain country of origin. Hence, no investigation of bilateral movements takes place which limits the results to the identification of push factors in the origin country.

Davenport et al. (2003) extend this study by including Internally Displaced People (IDP), thus examining the whole group of forced migrants, those who leave the country and those who stay within the borders. Additionally, they understand forced migration still as a decision people make, and therefore examine factors that push people out of their home country as well as factors that draw people to a country of destination. They implement this theoretical approach by using net migration stock as the dependent variable. Nevertheless, their findings are similar to those in Schmeidl (1997): measures of conflict, violence and political threat highly influence refugee movements while economic and population factors do not. Further, more autocratic governments impede refugees from leaving the country.

Yet another approach to study forced migration is taken up by Moore and Shellman (2004). Specifically, they are the first to study flows (compared to stocks) of refugees and IDPs, thereby examining push factors in the country of origin. The flow variable, that is constructed by subtracting stocks of subsequent years, reflects annual migrant movements. Again, conflict variables seem to be more important determinants of flight than economic or institutional ones. However, Moore and Shellman (2004) find that government and, especially, dissident violent behavior are the strongest predictors of refugee flows.

In their follow-up studies Moore and Shellman (2006, 2007) emphasize that fleeing is still a choice and analyze which factors determine the choice of destination forced migrants make. Although, distance is expected to play a key role as it increases transaction costs the overwhelmingly large share of forced migrants are refugees compared to only few IDPs. The comparison of refugees and IDPs reveals further that genocide predicts cross-border movements whereas civil war is a major determinant for internal displacements (Moore and Shellman, 2006). In the next step, they take a closer look on international movements and analyze why people seek refuge in some countries of destination rather than in others (Moore and Shellman, 2007). To do so, flows between country-dyads are examined with a special focus on cost and pull factors. Results suggest that people often

flee to bordering countries that themselves experience civil war or are even at war with the home country. These findings are consistent with Moorthy and Brathwaite (2016) who contend that states accept refugees from rival countries in order to undermine the enemy. Economic and institutional pull factors do not seem to play a large role whereas cost factors such as the size of the diaspora and a common border are highly significant predictors of international refugee migration.

In a similar fashion as Moore and Shellman (2006, 2007), Neumayer (2004, 2005) examines asylum migration to Western Europe. His results differ in that economic reasons are much more pronounced with the above mentioned determinants still being important drivers of flight and destination choice. The evidence suggests that the difference is due to the limitation of destination countries to Western Europe where most refugees have a long journey that requires preparation and accumulation of financial means and who are thus driven more by economic opportunities (Moore and Shellman, 2007).

Secondly, our study can contribute to the strand of literature that examines gender-specific migration patterns, where the large majority focuses on labor migration to developed countries. Among the first, Pedraza (1991) provides an extensive literature review on women in migration that describes determinants as well as consequences in origin and destination country of female migration.

As for empirical analyses, there are some studies that examine female migration to OECD countries mainly in the light of skill levels and discuss potential consequences for home and host countries (Dumont et al., 2007; Docquier et al., 2009, 2012). High-skilled female migration is rising compared to numbers of male counterparts (Dumont et al., 2007; Docquier et al., 2009), leading to a female brain drain that especially hits the poorest countries implying negative effects on health and education measures in the country of origin (Dumont et al., 2007). However, Docquier et al. (2012) report that, despite increasing female emigration rates, there is no genetic or social gender gap in high-skilled migration. They allocate the differences in rates to family reunification in the sense that mostly women tend to follow their husbands (rather than vice versa) as well as a heterogeneous response for men and women to traditional push factors - a finding that we verify in this study for the refugee population. When it comes to a differential gender-specific response to migration determinants, Beine and Salomone (2013) cannot support this for network externalities. They find that the effect of a diaspora is the same across gender.

Another nexus that is studied in the gender-sensitive migration literature is the one of gender inequality and female migration. The evidence on this relationship, however, is still inconclusive: Baudassé and Bazillier (2014) find that higher gender equality in the country of origin increases migration of, especially high-skilled, females. They find no support of the push hypothesis that low gender equality makes women leave the country. Ruysen and Salomone (2018) use individual level data to not only study actual female

migration but also women's intention to migrate. Their findings are somewhat in line with the above, as they report that the intention to leave the country increases with high levels of perceived gender discrimination (reflecting the push hypothesis) but the actual movement is much more dependent on 'traditional factors' such as financial situation and networks. Nejad and Young (2014) study the relationship for high-skilled females in a bilateral gravity setting. Their results suggest a non-linear behavior of the nexus: at high levels of the women's rights gap (meaning a large difference in women's right between destination and origin country) a decrease in the gap increases female migration. For small gaps, however, an increase in the gap raises the brain drain ratio of females compared to males.

Lastly, our paper is also related to studies looking into consequences of refugee inflows for destination countries such as destabilizing and conflict increasing effects by examining heterogeneous compositions of refugee inflows. The heterogeneity mainly looked at so far is ethnicity of refugee groups. As for now, the existing evidence is inconclusive on the specific question of the effect of refugees on conflict situations in host countries, see e.g. Böhmelt et al. (2019); Rügger (2019); Choi and Salehyan (2013); Salehyan (2008).

4 Refugees in migration theory

Since we argue that refugees are a subgroup of migrants we want to elaborate how existing migration theory can be applied to the specific group of refugees. Following the standard economic approach, migrants underlie a cost-benefit-calculus where they weigh the costs of leaving their home country against the expected benefit they receive from arriving in their country of destination (Sjaastad, 1962; Borjas, 1989). This calculus is dependent on personal as well as country specific characteristics. In the context of labor migration, factors such as (individual) skill level and (country-pair specific) wage differentials are usually major determinants of migration movements. Lee (1966) puts this calculus in the well-known framework of push and pull factors, intervening obstacles and personal factors. Push factors affect a person's choice to leave his or her home country such as unfavorable political or economic opportunities (Mayda, 2010; Meierrieks and Renner, 2017), discrimination (Ruyssen and Salomone, 2018), terrorism (Dreher et al., 2011) or conflict, persecution or political instability (Stanley, 1987; Schmeidl, 1997; Davenport et al., 2003; Neumayer, 2005; Echevarria and Gardezabal, 2016) as well as environmental circumstances (Beine and Parsons, 2015, 2017). Pull factors such as better economic conditions or higher labor outcome determine the country of destination (e.g. Borjas, 1987, 1989; Mayda, 2010; Grogger and Hanson, 2011; Beine et al., 2019)). Intervening obstacles or costs of migration may be country-pair characteristics like visa regulations, diaspora, distance and travel costs or cultural distance measures (Belot and Ederveen, 2012; Neumayer, 2015; Bertoli and Moraga, 2015; Czaika and Parsons, 2017; Krieger et al.,

2018; Adserà and Pytliková, 2015). Lastly, personal characteristics influence whether, when and where a person may migrate. Besides the economic and family situation, factors like ethnicity, age and gender (Docquier et al., 2009; Rüegger and Bohnet, 2018) can play an important role in the decision making. The personal characteristic most studied is the level of education (Borjas, 1987; Docquier and Marfouk, 2006; Grogger and Hanson, 2011; Bertoli et al., 2013).

Skill or education as a decisive personal characteristic is also what Grogger and Hanson (2011) put forward in their adaption of the Roy-Borjas-Sjaastad model to distinguish the dimensions i) scale of migration (i.e. how many people leave) ii) selection (i.e. who leaves) and iii) sorting (i.e. where do people go). Regarding selection, Grogger and Hanson (2011) follow Borjas' idea of self-selection⁴ into migration: Depending on personal characteristics (skills) individual's utility and costs of migration differ. This implies that the migrant population is not random, but depending on the skill-level of individuals. And from this selected group leaving their home country, individuals choose a destination – and again, this sorting is not random but depending on some selection criteria.

We argue the framework of push and pull factors in combination with a non-random selection process into migration is applicable to migrants even beyond migration for economic reasons. Combining it with other approaches to migration can enrich the economic point of view in a helpful manner. For example, Timmerman et al. (2014) as well as Kuhnt (2019) take on a different perspective on the determinants of migration with building a framework of macro-, meso- and micro-level factors. Macro-level factors represent factors like political institutions or national economic situation that apply equally to all migrants in a country. The meso-level reflects conditions within the local community or a comparable sub-national entity. It captures, for example, cultural and ethnic factors as well as networks and diaspora of migrants. Lastly, characteristics of the smallest unit of concern, the individual, are an important driver of migration. Gender, age or educational level of the individual migrant certainly shape his or her decision to move. Kuhnt (2019) also highlights the interactions that take place among factors of the various levels. Hence, a macro-level determinant like human rights violation may have a different effect on migration decisions for males and females—a relationship we are going to explore and empirically test in this study.

This emphasis on the interplay of macro-level drivers and individual characteristics is incorporated in other social sciences' literature on migration as well. Carling (2002) and later de Haas (2014) elaborate the idea of migration being a function of *migration aspiration* and *migration (cap)abilities*. Only for those persons that have the desire to migrate (aspiration) and at the same time are able to do so ((cap)ability) migration will be the outcome. Both decisive factors consist of the environmental context as well as

⁴Borjas (1987) combined the theory of migration as an investment decision by Sjaastad (1962) with the selection on individual characteristics for income maximization by Roy (1951).

personal characteristics (Carling and Schewel, 2018).

We now want to combine these theories of migration and apply it to refugee migration. Both, the cost-benefit-calculus of the Roy-Borjas-Sjaastad model and the concept of push and pull factors, have faced critique as they make strong assumptions on individuals having full information and making purely rational decisions. Further, it is claimed that they are not able to capture the complexity of the migration process and treat the migration decision as exogenous (Langley et al., 2016; de Haas, 2014). Nevertheless, we argue that the push-pull-model is a useful tool to structure the complex decision making process of all migrants as well as it enables a thorough empirical analysis.

In general, push factors are most relevant in a refugee's decision process (Schmeidl, 1997). In the case where people feel a threat to their lives this threat will most likely take up the major part of the decision to abandon their home. However, when analyzing refugee movements just looking at determinants in the home country neglects a relevant part. Existing research emphasizes that people - including refugees - have the choice to stay or to leave, and when opting for leave they have a choice of where to go (Richmond, 1993; Neumayer, 2005; Melander and Öberg, 2006; Davenport et al., 2003; Moore and Shellman, 2006; Iqbal, 2007; Krieger et al., 2020). This choice is reflected by factors that make an attractive destination country or that reduce the costs of the flight. Again, it is important to clarify this point: The more forced a migrant is to leave his or her home, the higher the relevance of push factors in the individual decision-making. Nevertheless, pull and cost factors play at least some role when people decide to leave and have to choose a country of destination. Hence, refugees can be described as "people making choices under highly constrained circumstances" (Moore and Shellman, 2007).

Nevertheless, not all refugees being exposed to the same situation make equal decisions whether, when and where to flee. In a village that suffers from frequent military attacks for instance, still some people decide to leave while others decide to stay as well as the timing and destination varies among those who actually flee although push factors such as exposure to violence as well as pull factors (e.g. economic or security situation in destination country) and cost factors (e.g. travel distance) are the same for all of them. Hence, we can observe heterogeneous responses to the same drivers of flight. In the standard labor migration model, such heterogeneity comes mainly from the skill level that individuals have and that influence their migration decision. For the decision to flee, we argue that gender is an important determinant for heterogeneity. We expect men and women to react differently to determinants of flight for mainly two reasons. First, people who share a common trait like gender or age have usually more similar preferences than they share with the 'out-group'. This is particularly true for the difference in preferences between men and women (Croson and Gneezy, 2009; Falk and Hermle, 2018). Second, vulnerability and possibility in home and destination country as well as on the route differ heavily for men and women.

In the following, we want to elaborate these two sources of heterogeneous effects along the three dimensions of push, pull and cost factors and derive hypotheses for our empirical investigation. This inspection helps us to shed a light on gender differences for the decision whether and when to flee (push factors) as well as where to go (pull and cost factors).

Push factors: Conflict and violence are found to be the major determinants that make people leave their home country (Schmeidl, 1997; Moore and Shellman, 2004; Iqbal, 2007; Echevarria and Gardeazabal, 2016). However, vulnerability of men and women in the face of a conflict situation may look different depending on the type of conflict. If the conflict is very intense all people including civilians are heavily affected and will seek for refuge. Hence, we expect to see no difference in flight patterns for men and women if conflict intensity is high. For conflicts with lower intensity, vulnerability in the home country may be higher for men. Buvinic et al. (2013) report numbers from different sources where male mortality exceeds female war-related deaths by far. One explanation is that participation of men in conflict is usually much higher. Most of the fighting force, be it the military or other rebel groups, are male which makes staying more risky for them compared to females. If civilians are largely unaffected by the conflict, for women the risk of staying is relatively lower.

Lastly, economic concerns may as well play a role in the decision who flees and when. As the UNDP (2019) reports for irregular migration to Europe, migrants with higher education receive substantially more financial support from their families and friends for their journey. This hints towards refugees being an investment for the (extended) family. In such cases, the household decides to send away the person that has the most possibilities abroad and thus can best support the family at home. Mostly, young male family members fulfill these requirements and are chosen (World Bank, 2017, p.49). However, we only expect such mechanism to play a role in the face of a low-intensity conflict as otherwise there is no time for such considerations.

H1: For low-intensity conflicts, risk of staying and possibility of leaving is higher for men than for women. This lowers the share of women in the refugee population compared to high-intensity conflicts.

If the dimension of the conflict extends and civilians are attacked and instrumentalized the risk calculation may look different. Sexual violence, for example, is a frequently used strategy, weapon or at least phenomenon in armed conflicts. Then, women are much more vulnerable as they are the prime victims (Koos, 2017; Cohen and Nordås, 2014). Since conflicts create a situation where especially the opportunity and the incentives for sexual attacks are much higher than in peacetime, females may be indirect victims of the fighting (Wood, 2006). The prevalence of sexual violence in a

conflict leads then to a higher risk for women to stay in the home country.

H2: For conflicts where civilians are instrumentalized, females are more likely to leave the country. This increases the share of women in the refugee population.

Cost factors: Concerning the costs of flight or migration, the distance to travel from home to destination country is a highly relevant factor (Lucas, 2001; Neumayer, 2004; Iqbal, 2007; Echevarria and Gardeazabal, 2016; UNHCR, 2019b). Besides the financial requirements a longer journey poses, the higher cost may be due to refugees on the move often being subject to violence, torture or slave-like treatment (UNHCR, 2019a; MixedMigrationCentre, 2018). The longer the journey the more opportunities for such danger occur. For females, traveling longer distances exhibits an exceptional risk as it means a higher exposure to sexual and gender-based violence (Freedman, 2016; UNHCR, 2019a). Not only are women regularly sexually assaulted by smugglers or security forces but also exposed to sexual attacks by other refugees on the journey or in the camps. Hence, distance as a cost factor of migration may account stronger negatively for female refugees' decision making than for male.

H3: Larger distances between home and destination country reduce the share of women among all refugees.

Another dyad-specific factor that influences migration flows is the existence and size of the diaspora in the destination country. Similar to the voluntary migration literature (see e.g. Beine et al., 2011; Beine and Salomone, 2013; Collier and Hoeffler, 2018) there is broad evidence that social networks increase forced migration (Moore and Shellman, 2007; Havinga and Böcker, 1999; Adhikari, 2012; Rüegger and Bohnet, 2018). Nevertheless, there are also findings for the opposite relationship (Melander and Öberg, 2006). Looking at gender differences, it is found that females are more sensitive towards (male) diaspora, however this is potentially driven by family reunification where wives follow their husbands (Docquier et al., 2012; Beine and Salomone, 2013; Cummings et al., 2015). Yet, other scholars argue that women value social networks higher and rely more on information of their network than men do (Docquier et al., 2009). Both arguments direct at diaspora having a stronger positive effect on females than on males. A similar line of argumentation may account for cultural factors having a heterogeneous effect on the choice of the destination country for men and women. Several measures that reflect a common culture (e.g. colonial history, common language, political or ethnic ties) between two countries have found to be relevant determinants of international migration movements as they ease information gathering, search costs, psychological and integration costs (Havinga and Böcker, 1999; Belot and Ederveen,

2012; Neumayer, 2004; Krieger et al., 2018; Rügger and Bohnet, 2018). While these studies do not differentiate migrants by gender, some of them find heterogeneous effects when differentiating migrants according to skill levels.

H4: Social networks in the destination country have a stronger positive effect on women's decision to flee. Thus, it increases the share of females in the refugee population.

Pull factors: Lastly, we turn to the sorting decision, that is the question, which destination an individual picks for asylum. More specifically, we aim to test whether there exist gender specific differences in such pull factors. A different treatment of men and women - usually to the disadvantage of females - can shape migration decisions differently for both genders. Gender discrimination in the country of origin has been found to influence female decision making to leave the country (Ruyssen and Salomone, 2018; Baudassé and Bazillier, 2014). Hence, we argue that gender inequality may have a differential impact as a pull factor as well. We expect that higher gender inequality in the host country lowers the share of female relatively to male refugees.

A similar reasoning may be valid for the presence of violence in potential countries of destination. As in their home country security issues are a major concern for most refugees, this may possibly be the case for potential destinations as well. Iqbal (2007) reports that war in the destination country decreases refugee flows. To the extent that women are more risk-averse and value security higher than men (Croson and Gneezy, 2009), this relationship may look different for both genders. Less violence and more security in the destination is thus expected to have a larger positive effect on female than on male refugees.

Moore and Shellman (2007) find that the economic situation in the destination country as well as the political system are not significant determinants for the sorting of refugees. Based on Meierrieks and Renner (2017) who find no differential effect for male and female migration with respect to economic freedom, we expect that political and economic factors do not influence the composition of refugees.

H5: Gender inequality and the presence of violence in the destination country have a stronger negative impact on female refugee flows than on male which lowers the share of women among all refugees.

Political and economic factors should not affect the composition.

5 Data

5.1 Demographic Data on Refugee Flows

The UNHCR provides time series on dyadic stocks of refugees since 1951, which are widely used for studies concerned with determinants and dynamics of refugee flows or effects of those flows (e.g. Moore and Shellman, 2006, 2007; ?; Echevarria and Gardeazabal, 2016; Dreher et al., 2019). The data comes from UNHCR offices who rely on their own data or get data from the government of the asylum country (UNHCR, 2017, p.3). To our knowledge, we are the first researchers who got access to dyadic stock data with demographic information through the UNHCR. The data is available for 2000–2015 and provides stocks of refugees by sex and age cohort from a given origin country living in a country of asylum in every December.⁵ Thus, we can assume that the data collections works in the same way as for the publicly available data: The UNHCR gathers data as it is supplied by governments in asylum countries. In some cases, the UNHCR itself is involved in dealing with refugees, so that data collection is even easier. As the data has not been released publicly, data cleaning was required and one interesting (and limiting) pattern is obvious. Many high-income countries do not report demographic data on the refugee population they host, a fact the UNHCR has criticized earlier (UNHCR, 2007, p. 47). Reasons for those data gaps could be differences in national data collection as well as definitions (e.g. Angenendt et al., 2016). In part, this can be attributed to the fact, that those countries do follow their own procedures and protocols as well as definitions and are less reliant on support by the UNHCR. Reasons for not publishing detailed demographic information on a regular basis may be politically motivated or organizational. In more practical terms, this implies that we are only able to analyze refugee flows with respect to gender for a subset of all countries for which the UNHCR reports data, which leads us to the first important step: Sample restrictions.

Sample restriction To avoid selection-into reporting effects, we conduct several sample cuts: First, we restrict our sample to all country-pairs for which the dataset contains at least one observation. Second, we only include countries in Africa and Asia, to avoid having only some countries in other parts of the world which may come with systematic differences as described above. Figure 4 shows the development of refugee stocks by continent of asylum for our period of observation. Notably, the y-axes require different

⁵The data is entitled “refugee population”, but without a detailed information on the legal categorization of the individuals as provided in the general UNHCR time series, where recognized refugees, asylum seekers and persons in refugee-like situations are distinguished. This makes comparison with studies using the publicly available time series without demographic information difficult. However, for the question at hand: What drives the composition of the refugee population living in a certain destination, this is of lesser importance. For questions such as work permits or integration classes these legal categorizations may be important. However, there has been little change for those questions so that our approach using fixed effects will capture most of this. As a consequence, we refer to ‘refugees’ throughout the text.

scaling as the size of stocks in Asia and Africa is a multiple of the size in Europe and America. Moreover, the graphs show difference in reporting. The “total” refers to the total stock (both genders and all age cohorts) in the demographic data. The “total, no demo” refers to the standard UNHCR time series data, where demographics are not reported. The differences in reported figures are smaller in Africa and Asia. By focusing on these two continents, we reduce reporting anomalies and at the same time include the most affected countries.

Figure 4 about here

Imputation Generally, the UNHCR time series as well as the demographic data comprises only county-pairs for which the refugee stock is positive (Marbach, 2018). If the country-pair is not reported, most studies assume that the stock of refugees is zero for this dyad (e.g. Dreher et al. (2019); Rüegger and Bohnet (2018); Echevarria and Gardeazabal (2016)). Marbach (2018) argues that this might be reasonable for the majority of not reported cases, but points out that this imputation disregards that there are often stocks of refugees where the origin country is coded unknown. Marbach (2018) tests two other imputation procedures, first linear interpolation using reported values from the respective dyad and second a least square prediction using information from other dyads. Both imputations outperform the simple imputation of zeros. For the moment, we follow a conservative version of the linear interpolation as we linearly interpolate only in cases where a country-pair time series has a one year gap. For all other missing observations per dyad, i.e. more than one in a row or if the first or the last year is missing, we do not interpolate. Instead, we follow the standard approach for those cases and impute a zero.

From stocks to flows to shares From the dyadic stock data, i.e. the number of refugees R of gender g (female or male) living in a country of asylum a and fleeing from their home country h we construct the gendered flow of refugees for year t :

$$R_{hat}^g - R_{hat-1}^g = Flow_{hat}^g \quad (1)$$

This corresponds to gross flows, as we cannot observe where the changes in stocks come from. Of course, the changes in stocks are composed of newly arrived refugees minus those who left (either returning or moving on) or died. While this procedure is far from optimal, it is the only feasible way to assess movements of refugees⁶. Whenever the gross flow turns negative, this implies that more people from country h have left compared to new arrivals in the country of asylum. We set these flows to zero as we are interested in

⁶For voluntary migration, many studies proceed similarly, as inflows (or outflows) of migrants are rarely published. In contrast, information regarding the stock of (e)migrants who are registered legally in the destination country is more easily to obtain.

positive gross flows only. As a robustness check we drop those flows (which changes the results only very little).⁷

As we are interested in what drives the gendered composition of refugee movements, we take the flow of refugees by gender from equation (1) and derive what we call the female 'share' in equation 2.

$$Share_{hat}^f = \frac{Flow_{hat}^f}{Flow_{hat}^f + Flow_{hat}^m} \quad (2)$$

The female share is our dependent variable of interest to answer the question what determines the composition of refugee flows. A share of 0.5 implies that a bilateral flow is balanced with respect to the gender composition, while below (above) 0.5 the flow consists of relatively more (less) women than men. In our underlying sample we observe 18.345 dyad-years⁸, with a mean share of 0.1153 (standard deviation is 0.2582). This highlights two important things: First, the composition of refugee flows is highly dominated by men. Second, the distribution is skewed to the left, indicating that OLS may not be suitable for the estimation (see section 6). If the share is increasing, this implies that the composition of a flow becomes relatively more female.

However, this compositional measure cannot show us, where changes in the composition come from. As equation 2 is a ratio from female flows over all flows in a dyad-year, it depicts only composite effects stemming from reactions in those flows. Any change in the composition is therefore depending on changes in the flow of female or male refugees and whether these changes are similar or differ in size and/or direction. Equation 3 tries to summarize these. For example, assuming that the composition of dyadic refugee flows becomes more female, this would be measurable in an increase in the share. The interpretation would be that relatively more women than men have moved from country h to a in year t . This effect can be either due to case (i) the flow of female refugees increases, but male flows do not change or (ii) the flow of female refugees does not change, but less men flee from h to a § or (iii) both flows increase, but female flows more strongly than male flows or (iv) both flows decrease, but male flows more strongly than female flows. Thus, to be able to tell where changes in the composition of refugee flows come from, we have to look into how the flow by gender (i.e. equation 1) react to changes in our explanatory variables.

$$Share_{hat}^f \uparrow \uparrow \uparrow \uparrow = \frac{Flow_{hat}^f \uparrow = \uparrow \downarrow}{Flow_{hat}^f \uparrow \uparrow \downarrow + Flow_{hat}^m = \downarrow \uparrow \downarrow} \quad (3)$$

⁷Another idea would be to include the absolute value of negative flows. However, this attributes the negative flows to the dyad $h-a$ without knowing whether these people have really returned to h or whether they moved on to another country or died. Thus, for the question what shapes inflows of refugees, we do not consider this as an appropriate procedure.

⁸At the moment, we have slightly different sample sizes from dimension to dimension as some variable cover only a subset of countries. Within each dimension the sample size is stable with exceptions described in more detail.

- **I:** $Flow_{hat}^f$ changes and $Flow_{hat}^m$ is not affected
- **II:** $Flow_{hat}^f$ is not affected and $Flow_{hat}^m$ changes
- **III:** $Flow_{hat}^f$ increases more strongly than $Flow_{hat}^m$
- **IV:** $Flow_{hat}^f$ decreases less strongly than $Flow_{hat}^m$

5.2 Explanatory variables

The structure of our data, a dyadic panel, allows us to employ different dimensions of fixed effects in order to separately disentangle (i) Push factors, i.e. factors in the home country driving people into fleeing; (ii) Cost factors, i.e. characteristics of the country-pair that influence how costly the movement from home to asylum country is; and (iii) Pull factors, i.e. factors in the destination country that shape the decision where to go. In the following, the variables we use for those three dimensions will be shortly described.

Push factors: Using the the UCDP Georeferenced Event Dataset (Sundberg and Melander, 2013; Croicu and Sundberg, 2017) we construct a variety of variables regarding the presence of conflict in the home country. It is important to note, that Syria is not included in the UCDP data and therefore not in our sample. The conflict events coded here (resulting from armed force, resulting in at least one direct death) are coded on daily and village level. We aggregate this to state-year level and construct a dummy for any kind of conflict⁹. To capture the intensity of conflict, we construct a count variable for the number of total *fatalities* in logs, including civilians, battle-deaths and unknown deaths¹⁰. Furthermore, we create a *duration* variable that counts the number of consecutive years with conflict in the home country. A binary variable for very *long and intense* conflicts is coded one for every conflict that continues for at least five years and has on average more than 25 fatalities per year.

In addition to conflict (intensity), we assess how conflict strategies affect the population by using the log of *civilian fatalities* and *battle-related fatalities*, also using the UCDP Georeferenced Event Dataset. Furthermore, sexual violence in conflict (i.e. whether one of the conflict parties is reported to actively conduct sexual violence that is related to the conflict) is taken from the Sexual Violence in Armed Conflict Data Project (SVAC) 2.0 by Cohen and Nordås (2014)¹¹ and in detail described by Cohen and Nordås (2014).

Political oppression is captured twofold, for one we include *freedom from political killings* (0 if there is evidence for systematic politically motivated killings approved by

⁹The UCDP data distinguishes three kinds of conflict, depending on the actors involved (state-based, non-state, as well as one-sided conflict)

¹⁰We use the category of best estimates (Croicu and Sundberg, 2017, p. 26).

¹¹The SVAC uses conflicts as coded by the UCDP/PRIO Armed Conflict Database and also considers the 5 years after a conflict ends. This implies that only conflicts with more than 25 battle-related deaths are considered for this coding.

the government, 4 if there are none) and *freedom of religion* (0 if there is no freedom but religious practices are outlawed, 4 if freedom of religion is secure, no discrimination against religious groups from public authorities) from the Varieties of Democracy Dataset (V-Dem) (Coppedge et al., 2019). Lastly, *freedom from torture* is included in some specifications (not all, due to high collinearity), also from V-Dem, defined as 0 for systematic torture up to 4 (non existent).

To account for more structural, potentially influential home country characteristics, we also include the *polity score* (Marshall et al., 2017) ranging from -10 (autocracy) to +10 (democracy) as well as a dummy coding political *instability* when a home country experienced a change in the polity score of three or more points (following Fearon and Laitin (2003)). Furthermore, *GDP per capita* and *population size* (from Head et al., 2010) are used in logs. Additionally, to observe whether extreme poverty constraints matter, we add *child mortality* (under five years old, deaths per 1000 births) from the World Development Indicators.

Lastly, we add the logged stock of internally displaced persons to capture internal flight patterns. Due to data limitations, this implies a smaller sample and is therefore done only as a robustness check. As there is no clear responsibility for internally displaced persons, data collection is rather limited. We use the figures reported by the UNHCR Population Statistics.

Costs To analyze differential effects of cost factors we distinguish geographic characteristics such as *contiguity* or the weighted¹² *distance between two countries* (in km, logged). For cultural and institutional similarities which may correlate with lower information costs or lower costs for integration in the destination, we use a variety of variables. For voluntary migration, numerous results show that cultural (dis-)similarities matter for the sorting into destination countries and that effects are heterogenous across education levels of migrants (e.g. Belot and Ederveen, 2012; Falck et al., 2017; Adserà and Pytliková, 2015; Krieger et al., 2018; Collier and Hoeffler, 2018) First, binary variables capture common *colonial ties* or a *common official language*. Data comes from Head et al. (2010). To look more into details, we add *religious distance*, *linguistic distance* and *genetic distance* between two countries. All variables are taken from Spolaore and Wacziarg (2016).

A potentially cost-reducing factor are networks in the country of asylum as Rügger and Bohnet (2018) show for ethnic groups. We code a variable for the *diaspora*, that is the mean stock of refugees from the origin living in the destination over the last three years. We also calculate this measure for males and for females separately. Additionally, we code a variable for the stock of *legally registered migrants* from the home country in the asylum country in the year 2000 from the Global Bilateral Migration Database by

¹²We use the population weighted distance provided by Head and Mayer (2014). This measure takes population agglomeration into account and not only the distance between the two capitals.

Özden et al. (2011) .

Pull factors For the decision where to go, characteristics of the country of asylum matter. As people flee from conflict and persecution, we include variables regarding the peacefulness and security of countries such as the logged number of conflict-related *fatalities* as well as the binary variable for long and intense conflicts as described above, using the UCDP Geo-referenced dataset. Similarly, *freedom from political killings*, *freedom of religion* and the *polity indicator* capturing the political system as well as *political instability* are included here as well, as described above. Furthermore, the variable *exclusion by social group* captures whether belonging to a social group determines access to services or spaces (1 if high exclusion is prevalent, 0 if none) from Varieties of Democracy (Coppedge et al., 2019). For women’s rights we distinguish freedom of movement (from 0, no freedom, to 1, free movement) and property rights for women (from 0, no property rights, to 1, secure property rights), also from Coppedge et al. (2019).

Again, structural variables are added here to account for their effect for the sorting into a asylum country. These are the same as above: *GDP per capita*, *population size* and *child mortality*.

6 Estimation strategy

Following our argument in Section 1, we run a gravity model testing for the (gender-specific) determinants of refugee flows. Similarly, Echevarria and Gardeazabal (2016) have applied a gravity approach on refugee migration, however they relied on total stocks. We stick somewhat more closely to the recent approach in migration research and analyze dyadic refugee flows which other migration scholars have done before (for refugee migration, e.g. Moore and Shellman, 2007) and is also the original idea of the gravity model in the trade literature (Silva and Teneyro, 2006). The idea to use flows as the dependent variable is also intuitive as we want to study the effect on refugee movements rather than on refugee stocks. To be able to analyze whether a particular variable has a differential effect on females or males from the flow variables we use the share of females among all refugees as the dependent variable (for a detailed description see section 5).

For our estimation we set up three different equations capturing the aforementioned push, cost and pull dimensions. This enables us to examine these factors in a controlled setting being able to include different sets of fixed effects for each of the three equations in order to exclude influence of omitted variables. The inclusion of fixed effects is widely used for voluntary migration whenever a time dimension is available, for example by Bertoli and Moraga (2015); Beine and Parsons (2015); Beine et al. (2019).

$$Share_{hat}^f = \alpha_0 + \beta_1(conflict_{ht-z}) + \beta_2(politics_{ht-z}) + \beta_2(economic_{ht-z}) + \gamma_h + \gamma_{at} + \gamma_{ah} + \epsilon_{hat} \quad (4)$$

$$Share_{hat}^f = \alpha_0 + \beta_1(distance_{ha}) + \beta_2(culture_{ha}) + \beta_2(diaspora_{ha}) + \gamma_{ht} + \gamma_{at} + \epsilon_{hat} \quad (5)$$

$$Share_{hat}^f = \alpha_0 + \beta_1(security_{at-1}) + \beta_2(women's_{at-1}) + \beta_2(economic_{at-1}) + \gamma_a + \gamma_{ht} + \gamma_{ha} + \epsilon_{hat} \quad (6)$$

Equation 4 shows the estimation for gender specific push effects of refugee migration. To analyze how push factors, such as conflict (intensity) or political and economic variables in the home country, we include home country fixed effects γ_h , destination-year fixed effects γ_{at} and country-pair fixed effects γ_{ah} . This leaves only variation in time-varying home country characteristics. We include push in year t , arguing that flight is a rather short-term decision, taking into account the situation at the moment. Furthermore, our dependent variable captures the change in refugee stocks from December in $t-1$ to December in year t . At the same time, we do not fear strong reverse causality as the outflow of refugee is probably not fueling conflict at home or has any immediate effects on the economic or political situation. We also test one- and two year lags.

The cost effects of flight and their impact on the gender composition of refugees are estimated in equation 5. To leave only variation in dyadic characteristics, origin-year fixed effects γ_{ht} and destination-year fixed effects γ_{at} are included. Explanatory variables are measures of distance, culture and social networks.

Lastly, equation 6 estimates how pull variables can influence the gender composition of refugee flows. To control for potential covariates, we include destination fixed effects γ_a , origin-year fixed effects γ_{ht} as well as dyad fixed effects γ_{ha} . Independent variables are time-varying destination specific characteristics such as security, women's rights and the economic situation in the previous year. Note, that for this dimension, we include all variables with a one-year-lag in order to account for information lags.

With the three equations above we estimate the effect of several indicators on the gender composition of refugee movements. As explained in section 5, we cannot observe what drives the measured change in the composition - whether it is due to a change in the female flows or in the male flows or in both of them. Thus, in a second step of our analysis we try to disentangle where the effect comes from. To do so, we reconstruct our panel dataset to a stacked dataset that allows us to estimate the effect of all covariates for male refugee flows and the differential effect for female refugee flows. This means that we duplicate all observations, set a dummy variable for females to one for every other observation and estimate a fully interacted model. Thus, by interacting the explanatory

variables with the females dummy, we are able to estimate the differential effect that a variable has on females compared to males. It is important to note that the dependent variable is not the female share anymore but the flow of male and female refugees respectively. Equation 7 shows the stacked model that corresponds to equation 4. For equation 6 and 5, we proceed accordingly. While the share ranges only between zero and one, refugee flow numbers can vary from zero to several thousands for a dyad-year. This makes the flow measure much less sensitive to minor changes in refugee movements.

$$Flow_{hat} = female_{at} [\alpha_0 + \beta_1(conflict_{ht-z}) + \beta_2(politics_{ht-z}) + \beta_2(economics_{ht-z})] + \gamma_h + \gamma_{at} + \gamma_{ah} + \epsilon_{hat} \quad (7)$$

Our preferred estimation is a Poisson pseudo-maximum-likelihood (PPML) estimator to explain the differential change in dyadic stocks, i.e. flows, by gender.¹³ The PPML estimator is suitable for distributions that are skewed and have large numbers of zeros. For example, Silva and Tenreyro (2006) show that this estimator outperforms log-linear OLS regressions and is more consistent even with a large number of zeros (Silva and Tenreyro, 2011). Furthermore, PPML estimators for gravity equations in trade (see e.g. Head and Mayer (2014)) are quite common and with increasing availability of dyadic migration data, this method is also more often used for migration (e.g. Beine and Parsons (2015); Bertoli and Moraga (2015); Czaika and Parsons (2017); Beine and Parsons (2017)) and specifically refugee migration (Dreher et al., 2019; Echevarria and Gardeazabal, 2016). To be more precise, we apply a multinomial PML estimator as described by Head and Mayer (2014, p.179) and a PPML approach for the analysis of differences in flows.¹⁴

While parts of the literature on refugee migration analyze flows using a two step model that assesses selection into positive refugee stocks, we do not follow this approach at the moment. The argument in favor of such models (e.g. hurdle models) as used for example by Rügger and Bohnet (2018); Moorthy and Brathwaite (2016); Moore and Shellman (2007) is that the process of flight follows two connected, but consecutive processes: First, there needs to be positive flow for a dyad. This is not true for all possible country-pairs as only some countries will accept refugees or some countries will never be picked as a country of asylum. Only if this hurdle is passed, positive flows or stocks of refugees can exist. In contrast, Echevarria and Gardeazabal (2016) argue that this is a process that cannot be disentangled properly and that PPML estimators by themselves can be applied for data with a large number of zeros. Furthermore, we are only working with a subset of countries that receive positive inflows of refugees at some point, so that we do not apply a two-step procedure.

¹³The estimator cannot accommodate negative flows. As explained above and similar to Bertoli and Moraga (2015) we set those flows to zero.

¹⁴Due to the high-dimensionality of our data (as $t = 15$ and $n_i = n_j$), we use the `ppmlhdfc` command for Stata, which is based on the `reghdfe` command, written and described by Guimarães and Portugal (2011) and `ppmlhdfc` by Correia et al. (2019).

7 Results

7.1 Push factors

Effects on the composition of flows Table 1 differentiates how conflict intensity, political oppression and conflict strategies affect the share of females in total refugee flows. Throughout all specifications, the share of females in the dyadic refugee flow is the dependent variable and home country fixed effects as well as dyad fixed effects and destination-year fixed effects are included. Thus, only time-varying variables for the home country enter the regression.

Starting with the most simple regression in column (1), only intensity of conflict (*fatalities*) and oppression (*political killings*) are included. However, as the risk of staying is probably not expanding linearly with fatalities but depend on the level, we include fatalities squared in column (2). Indeed, the linear term is positive and highly significant, indicating that with an increase in fatalities, refugee flows become more female. This effect, however, is reduced for high levels of fatalities, as indicated by the negative squared term. This non-linear effect remains significant throughout the entire table.

For political oppression, measured by freedom of political killings¹⁵, there is no effect on the composition of flows. Similarly, looking deeper into effects of long (duration) or very intense (and long) conflicts in column (4) and (5), we do not find a significant relation with the composition of flows. Turning to the political system, none of the variables is significant in the first half of the table, where we look at conflict intensity.

Most interestingly, conflict strategies matter, in line with hypothesis 2: If women face more risks of staying due to more civilian fatalities (column 6–8), the composition becomes more female. Again, the effect is non-linear. Battle-deaths do not have a significant effect on the share of women. If we include sexual violence as a conflict strategy, this is highly significant and positive: The share of women increases (column 7 to 10). This effect remains the same, even if we return to our more general measure of total fatalities (column 9) or add the stock of internally displaced persons (column 10). As the stock of internally displaced persons mostly serves as a control for internal flight, but its inclusion does not change the results much while decreasing sample size, we refer to column 9 as our preferred specification.¹⁶

Decision timing Testing our assumption of the decision timing (i.e. including all variables from the year t to explain the composition of refugee flows in year t), we include

¹⁵Freedom from torture is included only in column (8) as it is highly correlated with freedom from political killings. It does not change the results and is not significant, thus we disregard it.

¹⁶Including the dyadic diaspora is possible, even though some variation is lost due to the destination-year fixed effects. Furthermore, this variable is by definition highly endogenous. Thus, we add it separately, and it does not change the results for all other variables qualitatively (see Appendix table A-1 for a detailed table. Diaspora itself has a positive significant effect, indicating that a larger diaspora increases the share of women.)

the influence of conflict intensity in lags in table 2. If we include one and two-year lags simultaneously, column (3) shows that it is current as well as conflict in the previous year that matter, both indicating a non-linear effect (positive for small levels and negative for high levels) on the share.

Interestingly, if we add all variables with a one year lag (2), we can see that GDP per capita of the previous year has a highly significant effect, while the effects of all other variables remain the same: If GDP per capita in $t - 1$ increases, the share of females in t decreases. This result is robust to taking all other variables from year t and adding GDP per capita and conflict with one lag (column 4) or also sexual violence with a lag (column 5). Regardless of what drives these differences (male or female reactions), we may argue that average income has different effects on the propensity to flee for men and women, reducing the relative share of women. Table 3 shows that, lagged and current GDP is significant and negative for men, potentially reducing the need to flee or increasing resilience. When interacted with female, the difference is negative for current GDP, but positive for lagged GDP. The latter effect is also larger and may indicate that higher income in the past (i.e. for fleeing with a longer time horizon) may matter for female flows by reducing poverty barriers.

Where do compositional effects come from? Turning the question where these results come from, the stacked model, explaining the flow of refugees and testing for heterogeneous effects for men and women, in table 3 can give further insights. We only test our three preferred specifications from before, that is: first, all push factors in year t plus GDP per capita from the year before; second, fatalities also lagged one year and lastly civilian versus battle-related fatalities.

Throughout the three specification, the dummy variable ‘female’ is significant and negative, indicating that female flows are significantly smaller than male flows. As we have seen in table 1, fatalities have a significant positive effect on the female share, ‘feminizing’ refugee flows, but with a decreasing trend. In the stacked model, male flows increase with increasing fatalities, however, this increase becomes smaller for high numbers of fatalities. The difference to female flows can be seen in the interaction terms. Even though these are not significant, a Wald test for joint significance of the interaction and main effects rejects the null hypothesis of no significance at the 1% level. As pointed out before, small changes in flows can affect the share significantly, but for the stacked model of flows, changes need to be larger in order to show up significantly.¹⁷ Thus, both sexes seem to react rather similarly to an increase in fatalities. Adding fatalities from the year before, male flows are not affected significantly. The difference for female flows is again not significant, but the positive sign (for the linear term) fits to the findings for the share: Female flows seem

¹⁷The difference in significance when we move from shares to flows and stacked models also hints to the problem of making different models comparable. As we alter the dependent variable (and the structure of the data), asymptotic behavior may differ.

to react stronger to an increase in fatalities, but the difference is small. this finding also holds for civilian and battle-related deaths in column (3).

Another factor with high significance in the share model is the presence of sexual violence in conflict. The stacked model can show a more nuanced picture of where the strong positive effects on the female share stem from: For male flows the effect is not significant¹⁸. The coefficient for the difference between male and female flows is positive (also not significant), again, a Wald test indicates joint significance of main and interaction effects. Even though this difference is not significant, it is large enough to have compositional effects, which is in line with hypothesis 2: If women are instrumentalized in conflict, they are more likely to flee than men which alters the composition of refugee populations.

Lastly, GDP in the previous year affected the composition of flows and the stacked model shows that these may be due to male flows reacting negatively to past GDP, while the interaction term for female flows is positive, pointing to a smaller effect for women.

Push effects in a nutshell Combining the results from the share and the stacked flow model, our results hint to rather small gender differences in the reaction to factors associated with the risk of staying: If conflict intensity increases, both sexes flee. If at all, female flows react a bit stronger than male flows. The most pronounced differences is, not unexpectedly, found for sexual violence presence in conflicts: Women react stronger to its prevalence than men.

These results are largely robust to dropping negative flows instead of setting them to zero or to dropping transit (Morocco, Tunisia, Turkey, Mali and Algeria) countries (see Appendix). Also, using OLS instead of PPML for the composition of flows (i.e. the share model) gives similar results (available upon request).

7.2 Cost factors

Effects on the composition of flows Testing for differences in the effects of the costs of fleeing for men and women, as put forward by hypothesis three and four, we start with looking at compositional effects by including origin-time and destination-time fixed effects, which leaves us with the opportunity to analyze time-invariant dyadic factors.

Table 4 supports both hypotheses: Distance (in terms of kilometer) has a significant effect on the composition of flows, strongly reducing the share of females. This effect is highly non-linear (from column 2 onwards), indicating that small distances negatively affect the composition and for large distances the effect of distance on the female share is decreasing. For small distances, the binary variable for contiguity is important: Neighboring countries receive flows that are significantly more female than the flows to

¹⁸We do not want to over-interpret the slightly negative effect in column (2) as men being attracted to stay in case of sexual violence. The effect is very small and imprecisely measured so that we rather interpret it as being not decisive for male flows

non-neighboring countries. For cultural distance, the only (though only at the 10 % level) significant effect is found for a shared official language. If two countries share an official language, the share of female refugees is higher.

Again, diaspora is added separately (column 6 onwards) for endogeneity concerns. However, the results for the other variables mostly remain the same, while diaspora exhibits the expected result: positive and significant, i.e. increasing the share of females, as hypothesized. Interestingly, the effects seem to be more strongly for refugee diasporas. The diaspora from legally registered immigrants loses significance if refugee diasporas are included (column 6 onwards). It does not matter whether diasporas are distinguished by gender (column 8 and 9).

Where do compositional effects come from? Looking into the different potential explanations for a change in the share of females, we turn now to table 5, column (1). Interestingly, this partly moderates previous results. The effect of linear distance is almost similar, i.e. both flows are reduced. While the squared term is not significant for men but negative, the difference for women is significantly negative. This result seems to contradict the effect on the share, however, we do not want to overinterpret it as the coefficient is rather small and only significant at the 10% level. The differences for small distances are most pronounced for countries sharing a border: Contiguity is positively and significantly affect male flows, and the effect for women is significantly stronger.

For cultural distances, the effects are mixed and only partly fit to the previous results. A shared official language is not significant for males nor are women reacting differently, but the interaction effect is positive, pointing to a stronger positive effect for women than for men. Interestingly, the interaction term for colonial ties is significant and negative for women. In combination with the positive effect for male flows, we may interpret this as women reacting in a different manner (with smaller flows) if a country-pair share colonial ties. Looking back into table 4, the comparable coefficient for colonial ties in column (5) is very small and imprecisely measured.

In column (2) and (3) female and male diasporas are added and the results support our hypothesis: If the refugee diaspora increases (regardless of whether it is male or female), male and female flows react similarly, but the interaction effect (i.e. the difference between them) is significant negative as well, indicating that women react less strongly (potentially even in the other direction). Interestingly, the effect is negative, which is counter-intuitive and may be attributed to endogeneity concerns raised above.

Dyadic factors in a nutshell Again, we take the insights from the share and the stacked flow models together and can deduct that there is support for our hypothesis three: Neighboring countries are significantly more attractive for women and the composition of flows is reacting strongly as well. For long distances gender differences are significantly

different, with women react less appalled than men. Lastly, cultural differences affect both men and women rather similarly, with the exception of colonial ties or a shared official language. While for the latter, the difference is negative it is positive for a shared language, pointing to different needs of men and women.

As before, the results for non-negative flows only are shown in the appendix. Results change only slightly. Also, using OLS instead of PPML for the composition of flows (i.e. the share model) gives similar results (available upon request).

7.3 Pull factors

Effects on the composition of flows Lastly, we take the perspective of receiving countries, that is looking into drivers for the sorting into countries of asylum whether there are gender-specific effects. Including origin-year, pair- and destination fixed effects, we focus on time-varying factors associated with the attractiveness of a country of asylum. Table 6 starts by including (step-by-step) factors related to peacefulness and security in the destination, variables describing the political system and women’s rights and eventually economic factors. Column (5) is the most complete model and shows that for many factors no significant effect for the composition of refugee flows can be found. Peace in the destination seems to attract relatively more females. As expected, women’s rights are positively and significantly correlated with the female share.¹⁹ Column (6) adds the refugee diaspora (averaged over the last three years), which is, again, subject to endogeneity concerns, but does not change the results of any other variable strongly. Diaspora itself is negatively associated with the share of females in bilateral refugee flows.

Before we turn to the underlying reactions in terms of flows, we want to come back to the importance of neighboring countries. As the results for dyadic factors have already shown, this distinction is of high importance. Thus, table 7 tests for differences between neighboring and non-neighboring countries. In column (1) and (2), the previous specifications are interacted with a dummy for contiguity (column 2 is again including diaspora as well without a significant effect and not changing any other coefficient). To ease interpretation later on in the stacked flow model, we split the sample so that column (3) only looks at sorting into neighboring and column (4) into non-neighboring countries. What stands out is that peace duration is significantly changing the composition of in-flows for non-neighboring countries, but that this is not the case for neighboring countries. Similarly, freedom of movement for women is significantly positively associated with the share of females for non-neighboring countries, but for neighboring countries, this effect is smaller. The exclusion by social group index matters strongly for the composition of flows (decreasing the female flows relatively more) into neighboring countries, but less for non-

¹⁹The results of column (5) do not change when we exclude the most dominant transit countries for refugee movements towards the European and American continent. These are Morocco, Tunisia, Turkey, Mali and Algeria (see Appendix) .

neighboring countries. Furthermore, political instability is negatively associated with the female share (though not significant) for non-neighboring countries but for neighboring countries this is even more pronounced. These effects are all confirmed by sample splits. The mortality rate of children is only significantly altering the share for non-neighboring countries, while fatalities are only significant for the composition of flows into neighboring countries. Lastly, the size of the diaspora is only significantly associated with the share of women fleeing into neighboring countries.

Thus, we can conclude already, that there are manifold factors in destination countries associated with the composition of inflows and that it is important to distinguish neighboring and non-neighboring countries.

Where do compositional effects come from? Looking into the different potential explanations for a change in the share of females, we turn now to table 8. The full model is now depicted in column (1), while column (2) and (3) show the sample splits for neighbors and non-neighboring destinations respectively.²⁰

We first look into the dynamics of neighboring countries. In line with the findings for share, we find that fatalities deter inflows into neighboring countries while peace duration has no significant effect. Both effects are rather similarly for men and women. Political instability is significantly deterring male flows, while the interaction term is not significant it is nevertheless pointing to the same direction, i.e. a more pronounced effect for women, lowering their share. In the share model, exclusion by social group is significantly negatively associated with the share of women in the flow to neighboring countries. The results here show that for both male and the difference to female flows, the effect is very imprecisely measured as the standard errors are very high, probably driven by small variation in the indicator in this small sample. The effect for the size of the diaspora is negative for male flows²¹ and the difference is not significant and small for female flows. For the share, the effect was significant and negative. Again, the difference may be in the sensitivity towards small changes in the size of flows. Thus, for neighboring countries, the factors attracting refugees are rather similar across sexes: peacefulness and stability are decisive.

For non-neighboring countries, there are more differences: A longer duration of peace is associated with a higher inflow of men, while fatalities are not significant. For female flows, the difference for the effect of peace duration is not significant. The exclusion of social groups is significantly negative for male flows and the difference for women is not significant (though pointing to the same direction), supporting the findings for the composition. Political instability is not significant for men, though the sign of the effect is negative. For female flows, this deterring effect is significantly higher. Property rights

²⁰The sample split is more convenient than interpreting a threefold fully interacted model.

²¹This finding is partly contradictory to previous studies such as Rügger and Bohnet (2018), however, we use a rather simple measure for the size of the diaspora.

of women is positive and significant for men and again, the effect is stronger for women.²² GDP per capita is not significantly associated with changes in the male flows but the interaction effect shows that women are more deterred by it than men.

Sorting factors in a nutshell The question of where to go seems to be the most gender-sensitive part of the decision to flee. However, the distinction between neighboring and non-neighboring countries is crucial as well. For flows into neighboring countries, peace and stability are similarly decisive for men and women. The compositional effect of exclusion by groups is not supported by the stacked model. Thus, we may argue that neighboring countries serve as a next safe haven and that other factors such as the political and economic situation do not matter much, neither for men nor for women. For non-neighboring countries, the history of peace and the political system matter more strongly for both genders and if at all, men are more attracted by longer peace duration. Female flows are more sensitive to instability and women’s rights and the economic situation in the destination country.

Also, using OLS instead of PPML for the composition of flows (i.e. the share model) or dropping non-negative flows gives similar results (available upon request).

8 Conclusion

In this study we analyze gender-specific differences in the responses to a variety of flight determinants. Within the three dimensions of push, pull and cost factors, we examine which factors have heterogeneous effects on the decision to flee for men and women. For the empirical analysis, we use a novel dataset with dyadic refugee stocks including information on gender provided by the UNHCR for the years 2000-2015. Our results suggest that within the push dimension there is only little difference in gender-specific responses. Most prominently, women react relatively stronger to the existence of sexual violence as a conflict strategy. When it comes to dyadic factors, we find support for our hypothesis that distance is a decisive factor with men and women being affected differently. Neighboring countries attract more women – in comparison to men as well as to non-neighbors. The most gender-sensitive dimension appears to be the pull factors. Our results suggest that males and females sort differently into countries of destination. For flows into neighboring countries, most factors are similarly relevant for both genders. When fleeing to non-neighboring countries, female flows are more sensitive to political instability, women’s rights and the economic situation in the country.

So far, this study is a first exploration of potential gender-specific effects of drivers of refugee migration. Consistent with the theory (and the definition of a refugee), most

²²Interestingly, this pattern deviates from the share model. There, freedom of movement is significant while property rights is not.

people are forced to flee from violence which makes further push factors largely irrelevant. However, much more interesting is, how their decision of timing as well as the choice of their country of asylum is shaped. The consequence of a skewed gender composition of refugee flows may be a changing demographic structure in the origin or destination country. Such changes may have implications for fertility or on the labor market. Concerning integration policies or assistance in refugee camps, the composition of inflows may also be of interest for decision-makers and international organizations.

References

- Adhikari, P. (2012). The plight of the forgotten ones: Civil war and forced migration. *International Studies Quarterly*, 56(3):590–606.
- Adserà, A. and Pytliková, M. (2015). The Role of Language in Shaping International Migration. *The Economic Journal*, 125(586):F49–F81.
- Angenendt, S., Kipp, D., and Koch, A. (2016). Many refugees, poor data: Development cooperation requires higher-quality data. *SWP Comments*.
- Baudassé, T. and Bazillier, R. (2014). Gender inequality and emigration: Push factor or selection process? *International Economics*, 139:19–47.
- Beine, M., Bourgeon, P., and Bricongne, J.-C. (2019). Aggregate fluctuations and international migration. *The Scandinavian Journal of Economics*, 121(1):117–152.
- Beine, M., Docquier, F., and Özden, Ç. (2011). Diasporas. *Journal of Development Economics*, 95(1):30–41.
- Beine, M. and Parsons, C. (2015). Climatic factors as determinants of international migration. *The Scandinavian Journal of Economics*, 117(2):723–767.
- Beine, M. and Parsons, C. R. (2017). Climatic factors as determinants of international migration: Redux. *CESifo Economic Studies*, 63(4):386–402.
- Beine, M. and Salomone, S. (2013). Network effects in international migration: Education versus gender. *The Scandinavian Journal of Economics*, 115(2):354–380.
- Belot, M. and Ederveen, S. (2012). Cultural barriers in migration between OECD countries. *Journal of Population Economics*, 25(3):1077–1105.
- Bertoli, S., Fernández-Huertas Moraga, J., and Ortega, F. (2013). Crossing the Border: Self-Selection, Earnings and Individual Migration Decisions. *Journal of Development Economics*, 101:75–91.
- Bertoli, S. and Moraga, J. F.-H. (2015). The size of the cliff at the border. *Regional Science and Urban Economics*, 51:1–6.
- Böhmelt, T., Bove, V., and Gleditsch, K. S. (2019). Blame the victims? Refugees, state capacity, and non-state actor violence. *Journal of Peace Research*, 56(1):73–87.
- Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review*, 77(4):531–553.

- Borjas, G. J. (1989). Economic theory and international migration. *International Migration Review*, 23:457–485.
- Borjas, G. J. and Monras, J. (2017). The labour market consequences of refugee supply shocks. *Economic Policy*, 32(91):361–413.
- Buvinic, M., Das Gupta, M., Casabonne, U., and Verwimp, P. (2013). Violent conflict and gender inequality: An overview. *The World Bank Research Observer*, 28.
- Carling, J. (2002). Migration in the age of involuntary immobility: Theoretical reflections and Cape Verdean experiences. *Journal of Ethnic and Migration Studies*, 28(1):5–42.
- Carling, J. (2015). Refugees are also migrants. All migrants matter. Border Criminologies Blog. Oxford: Department of Criminology (<http://bordercriminologies.law.ox.ac.uk/refugees-are-also-migrants>, accessed 31/01/2020).
- Carling, J. (2017). Refugee advocacy and the meaning of ‘migrants’. *PRIO Policy Brief*, 2.
- Carling, J. and Schewel, K. (2018). Revisiting aspiration and ability in international migration. *Journal of Ethnic and Migration Studies*, 44(6):945–963.
- Choi, S.-W. and Salehyan, I. (2013). No good deed goes unpunished: Refugees, humanitarian aid, and terrorism. *Conflict Management and Peace Science*, 30(1):53–75.
- Cohen, D. K. and Nordås, R. (2014). Sexual violence in armed conflict. *Journal of Peace Research*, 51(3):418–428.
- Cohen, D. K. and Nordås, R. (2014). Sexual violence in armed conflict dataset. online, retrieved in December 2019.
- Collier, P. and Hoeffler, A. (2018). Migration, diasporas and culture: An empirical investigation. *Kyklos*, 71(1):86–109.
- Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., Bernhard, M., Fish, M. S., Glynn, A., Hicken, A., et al. (2019). V-dem [country-year/country-date] dataset v9.
- Correia, S., Guimarães, P., and Zylkin, T. (2019). ppmlhdf: Fast poisson estimation with high-dimensional fixed effects.
- Croicu, M. and Sundberg, R. (2017). UCDP GED codebook version 17.1. *Department of Peace and Conflict Research, Uppsala University*.

- Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–74.
- Cummings, C., Pacitto, J., Lauro, D., and Foresti, M. (2015). Why people move: Understanding the drivers and trends of migration to Europe. *London: Overseas Development Institute*.
- Czaika, M. and Parsons, C. R. (2017). The gravity of high-skilled migration policies. *Demography*, 54(2):603–630.
- Davenport, C., Moore, W., and Poe, S. (2003). Sometimes you just have to leave: Domestic threats and forced migration, 1964-1989. *International Interactions*, 29(1):27–55.
- de Haas, H. (2014). Migration theory: Quo vadis? IMI Working Papers 100, International Migration Institute, University of Oxford, Oxford.
- Docquier, F., Lowell, B. L., and Marfouk, A. (2009). A gendered assessment of highly skilled emigration. *Population and Development Review*, 35(2):297–321.
- Docquier, F. and Marfouk, A. (2006). International Migration by Educational Attainment (1990-2000) - Release 1.1. In $\tilde{\text{A}}\hat{\text{A}}_j\hat{\text{A}}_{\frac{1}{2}}\text{zden}$, C. and Schiff, M., editors, *International Migration, Remittances and Development*. Palgrave Macmillan: New York.
- Docquier, F., Marfouk, A., Salomone, S., and Sekkat, K. (2012). Are skilled women more migratory than skilled men? *World Development*, 40(2):251–265.
- Docquier, F. and Rapoport, H. (2012). Globalization, brain drain, and development. *Journal of Economic Literature*, 50(3):681–730.
- Dreher, A., Fuchs, A., and Langlotz, S. (2019). The effects of foreign aid on refugee flows. *European Economic Review*, 112:127–147.
- Dreher, A., Krieger, T., and Meierrieks, D. (2011). Hit and (They Will) Run: The Impact of Terrorism on Migration. *Economics Letters*, 113(1):42–46.
- Dumont, J.-C., Martin, J. P., and Spielvogel, G. (2007). Women on the move: The neglected gender dimension of the brain drain.
- Dustmann, C., Fasani, F., Frattini, T., Minale, L., and Schönberg, U. (2017). On the economics and politics of refugee migration. *Economic Policy*, 32(91):497–550.
- Echevarria, J. and Gardeazabal, J. (2016). Refugee gravitation. *Public Choice*, 169(3-4):269–292.
- Falck, O., Lameli, A., and Ruhose, J. (2017). Cultural Biases in Migration: Estimating Non-Monetary Migration Costs. *Papers in Regional Science*, forthcoming.

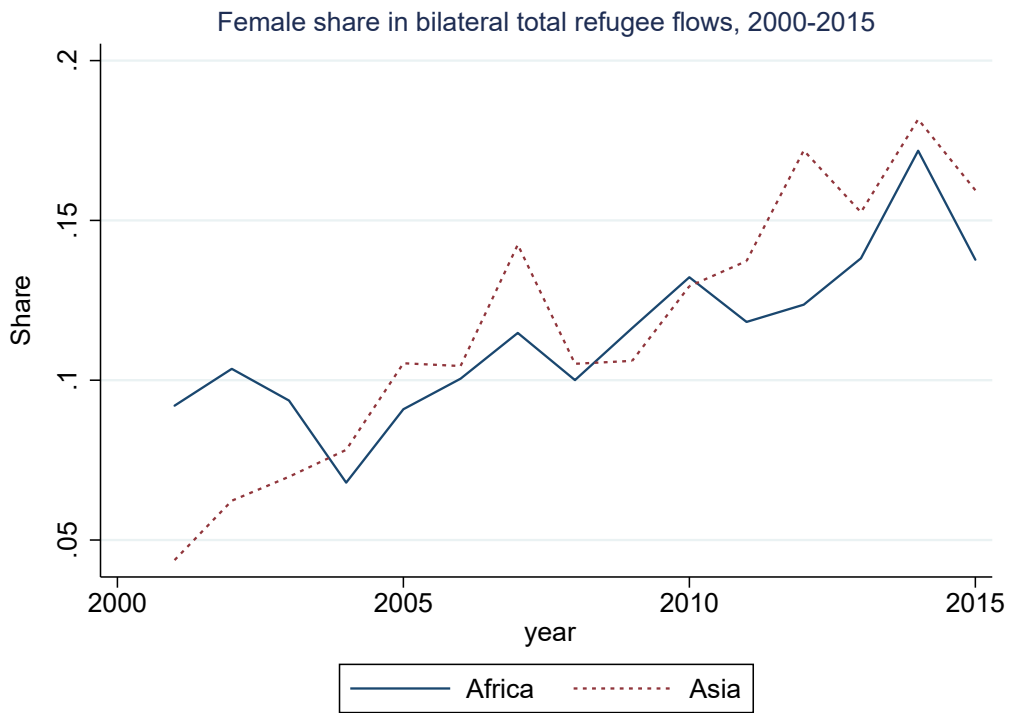
- Falk, A. and Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science*, 362(6412).
- Fearon, J. D. and Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American Political Science Review*, 97(1):75–90.
- Freedman, J. (2016). Sexual and gender-based violence against refugee women: A hidden aspect of the refugee "crisis". *Reproductive Health Matters*, 24(47):18–26.
- Gineste, C. and Savun, B. (2019). Introducing posvar: A dataset on refugee-related violence. *Journal of Peace Research*, 56(1):134–145.
- Grogger, J. and Hanson, G. H. (2011). Income Maximization and the Selection and Sorting of International Migrants. *Journal of Development Economics*, 95(1):42–57.
- Guimarães, P. and Portugal, P. (2011). A simple feasible procedure to fit models with high-dimensional fixed effects. *The Stata Journal*, 10(4):628–649.
- Havinga, T. and Böcker, A. (1999). Country of asylum by choice or by chance: Asylum-seekers in Belgium, the Netherlands and the UK. *Journal of Ethnic and Migration Studies*, 25(1):43–61.
- Head, K. and Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of International Economics*, volume 4, pages 131–195. Elsevier.
- Head, K., Mayer, T., and Ries, J. (2010). The erosion of colonial trade linkages after independence. *Journal of International Economics*, 81(1):1–14.
- Hudson, V. M. and den Boer, A. M. (2004). *Bare Branches. The Security Implications of Asia's Surplus Male Population*. The MIT Press.
- IOM (2019). Glossary on migration. International Migration Law 34, International Organisation for Migration, Geneva.
- Iqbal, Z. (2007). The geo-politics of forced migration in Africa, 1992—2001. *Conflict Management and Peace Science*, 24(2):105–119.
- Koos, C. (2017). Sexual violence in armed conflicts: research progress and remaining gaps. *Third World Quarterly*, 38(9):1935–1951.
- Krieger, T., Renner, L., and Ruhose, J. (2018). Long-term relatedness between countries and international migrant selection. *Journal of International Economics*, 113:35 – 54.
- Krieger, T., Renner, L., and Schmid, L. (2020). The individual level: Sorting effects. In Krieger, T., Panke, D., and Pregernig, M., editors, *Environmental Conflicts, Migration and Governance*, chapter 6, pages 103–120. Bristol University Press.

- Kuhnt, J. (2019). Literature review: drivers of migration. Why do people leave their homes? Is there an easy answer? A structured overview of migratory determinants. Discussion Paper 9/2019, German Development Institute (DIE), Bonn.
- Langley, S., Vanore, M., Siegel, M., Roosen, I., Rango, M., Leonardelli, I., and Laczko, F. (2016). The push and pull factors of asylum related migration: A literature review.
- Lee, E. S. (1966). A theory of migration. *Demography*, 3(1):47–57.
- Lucas, R. E. (2001). The effects of proximity and transportation on developing country population migrations. *Journal of Economic Geography*, 1(3):323–339.
- Marbach, M. (2018). On imputing UNHCR data. *Research & Politics*, 5(4):2053168018803239.
- Marshall, M. G., Gurr, T. R., and Jaggers, K. (2017). Polity IV project: Political regime characteristics and transitions, 1800-2017.
- Mayda, A. M. (2010). International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows. *Journal of Population Economics*, 23(4):1249–1274.
- Meierrieks, D. and Renner, L. (2017). Stymied Ambition: Does a Lack of Economic Freedom Lead to Migration? *Journal of Population Economics*, 30(3):977–1005.
- Melander, E. and Öberg, M. (2006). Time to go? duration dependence in forced migration. *International Interactions*, 32(2):129–152.
- MixedMigrationCentre (2018). Protection concerns of people on the move across west africa and libya. Technical report.
- Moore, W. H. and Shellman, S. M. (2004). Fear of persecution. Forced migration, 1952-1995. *Journal of Conflict Resolution*, 48(5):723–745.
- Moore, W. H. and Shellman, S. M. (2006). Refugee or internally displaced person? To where should one flee? *Comparative Political Studies*, 39(5):599–622.
- Moore, W. H. and Shellman, S. M. (2007). Whither will they go? A global study of refugees' destinations, 1965–1995. *International Studies Quarterly*, 51(4):811–834.
- Moorthy, S. and Brathwaite, R. (2016). Refugees and rivals: The international dynamics of refugee flows. *Conflict Management and Peace Science*, 36(2):131–148.
- Nejad, M. N. and Young, A. T. (2014). Female brain drains and women's rights gaps: A gravity model analysis of bilateral migration flows. IZA Discussion Papers 8067, Institute of Labor Economics (IZA).

- Neumayer, E. (2004). Asylum destination choice: what makes some west european countries more attractive than others? *European Union Politics*, 5(2):155–180.
- Neumayer, E. (2005). Bogus Refugees? The Determinants of Asylum Migration to Western Europe. *International Studies Quarterly*, 49(3):389–409.
- Neumayer, F. B. E. (2015). Spatial Dependence in Asylum Migration. *Journal of Ethnic and Migration Studies*, 41(7):1131–1151.
- Özden, Ç., Parsons, C. R., Schiff, M., and Walmsley, T. L. (2011). Where on earth is everybody? The evolution of global bilateral migration 1960–2000. *The World Bank Economic Review*, 25(1):12–56.
- Pedraza, S. (1991). Women and migration: The social consequences of gender. *Annual Review of Sociology*, 17:303–325.
- Richmond, A. H. (1993). Reactive migration: Sociological perspectives on refugee movements. *Journal of refugee Studies*, 6(1):7–24.
- Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2):135–146.
- Rüegger, S. (2019). Refugees, ethnic power relations, and civil conflict in the country of asylum. *Journal of Peace Research*, 56(1):42–57.
- Rüegger, S. and Bohnet, H. (2018). The Ethnicity of Refugees (ER): A new dataset for understanding flight patterns. *Conflict Management and Peace Science*, 35(1):65–88.
- Ruiz, I. and Vargas-Silva, C. (2015). The labor market impacts of forced migration. *American Economic Review*, 105(5):581–86.
- Ruyssen, I. and Salomone, S. (2018). Female migration: A way out of discrimination? *Journal of Development Economics*, 130:224–241.
- Salehyan, I. (2008). The externalities of civil strife: Refugees as a source of international conflict. *American Journal of Political Science*, 52(4):787–801.
- Salehyan, I. and Gleditsch, K. S. (2006). Refugees and the spread of civil war. *International organization*, 60(2):335–366.
- Schmeidl, S. (1997). Exploring the causes of forced migration: A pooled time-series analysis, 1971-1990. *Social Science Quarterly*, 78(2):284–308.
- Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 88(4):641–658.

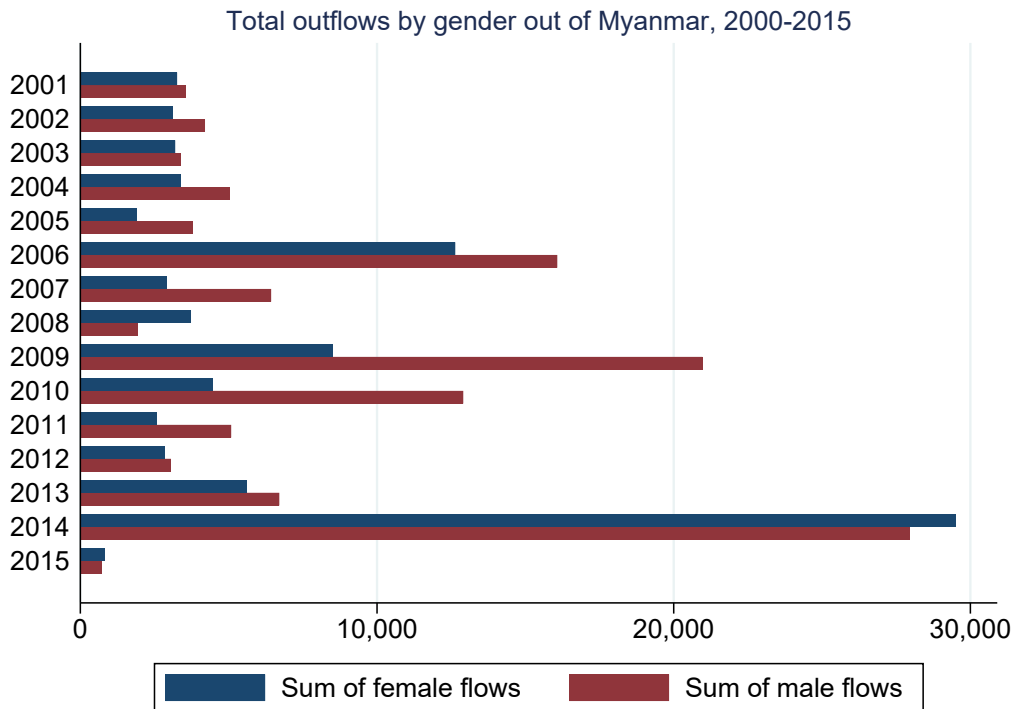
- Silva, J. S. and Tenreyro, S. (2011). Further simulation evidence on the performance of the poisson pseudo-maximum likelihood estimator. *Economics Letters*, 112(2):220–222.
- Sjaastad, L. A. (1962). The Costs and Returns of Human Migration. *Journal of Political Economy*, 70(5):80–93.
- Spolaore, E. and Wacziarg, R. (2016). Ancestry, Language, and Culture. In Ginsburgh, V. and Weber, S., editors, *The Palgrave Handbook of Economics and Language*, chapter 6. London: Palgrave Macmillan.
- Stanley, W. D. (1987). Economic migrants or refugees from violence? A time-series analysis of Salvadoran migration to the United States. *Latin American Research Review*, 22(1):132–154.
- Sundberg, R. and Melander, E. (2013). Introducing the UCDP georeferenced event dataset. *Journal of Peace Research*, 50(4):523–532.
- Timmerman, C., De Clerck, H. M.-L., Hemmerechts, K., and Willems, R. (2014). Imagining Europe from the outside: The role of perceptions of human rights in Europe in migration aspirations in Turkey, Morocco, Senegal and Ukraine. In *Communicating Europe in Times of Crisis*, pages 220–247. Springer.
- Tumen, S. (2016). The economic impact of Syrian refugees on host countries: Quasi-experimental evidence from Turkey. *American Economic Review*, 106(5):456–60.
- UNDP (2019). Scaling fences: Voices of irregular african migrants. Technical report, Regional Bureau for Africa, United Nations Development Programme, New York.
- UNHCR (2007). UNHCR Statistical Yearbook 2005.
- UNHCR (2017). UNHCR Statistical Yearbook 2016.
- UNHCR (2019a). Desperate journeys. refugees and migrants arriving in europe and at europe’s borders (january - december 2018). Technical report, UNHCR.
- UNHCR (2019b). Global trends forced displacement in 2018. Technical report, UNHCR, Geneva.
- UNHCR (2019). UNHCR Population Statistics Database. Available from http://popstats.unhcr.org/en/time_series, last accessed 1 February 2019.
- Wood, E. J. (2006). Variation in sexual violence during war. *Politics & Society*, 34(3):307–342.
- World Bank (2017). Forcibly displaced: Toward a development approach supporting refugees, the internally displaced, and their hosts. Washington, DC: World Bank.

Figure 1



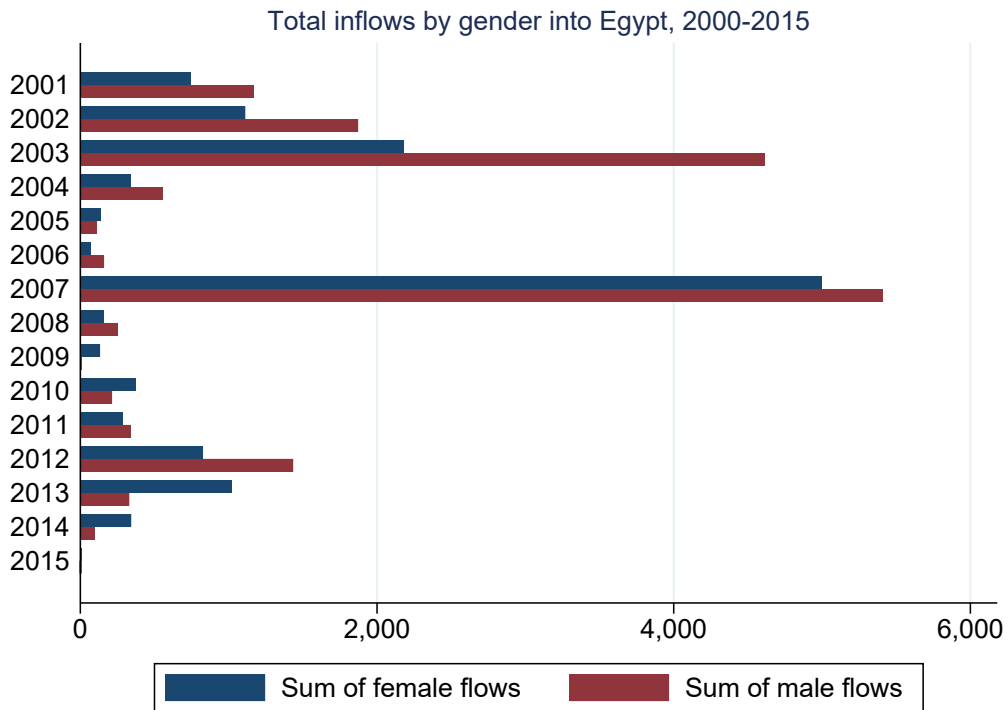
Source: Own calculations, data provided by UNHCR

Figure 2



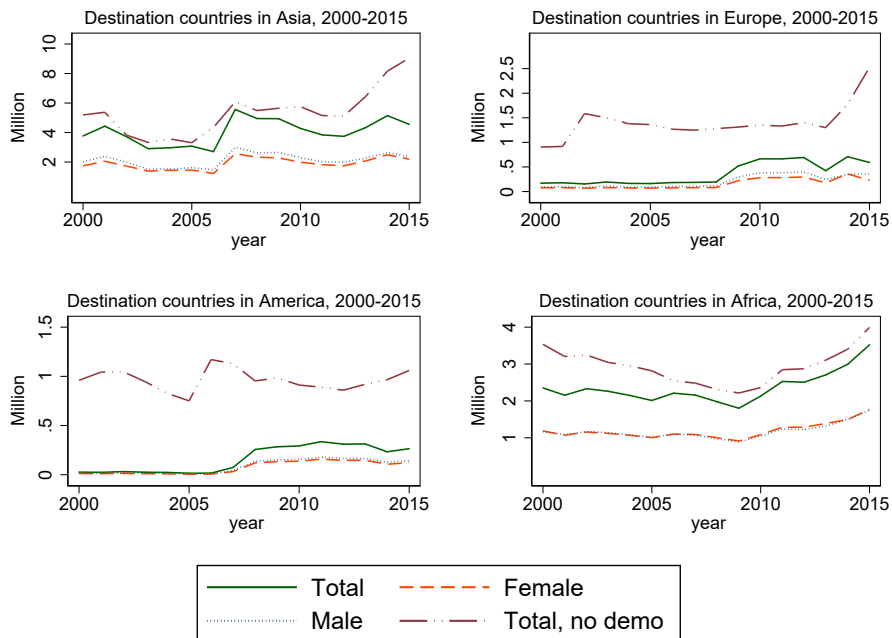
Source: Own calculations, data provided by UNHCR

Figure 3



Source: Own calculations, data provided by UNHCR

Figure 4: Refugee stocks by country of asylum



Notes: The figure shows how the sum of refugee stocks by continent developed between 2000-2015. The line “total” comes from the UNHCR data without demographic information. The deviations between this total and the total from the demographic data is largest for America and Europe. Notably, the increase in 2015 in Europe is missing for the demographic data.

Table 1: *Push factors and composition of flows*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Conflict and persecution					Conflict Strategies				
Fatalities (/1000)	0.020*	0.144***	0.126***	0.119***	0.119***				0.117***	0.110***
	(0.012)	(0.036)	(0.033)	(0.033)	(0.033)				(0.031)	(0.032)
Fatalities (/1000), squared		-0.009***	-0.008***	-0.007***	-0.007***				-0.007***	-0.007***
		(0.003)	(0.002)	(0.002)	(0.002)				(0.002)	(0.002)
Freedom from political killings	-0.070*	-0.030	-0.012	0.001	0.007	-0.010	0.015	0.020	0.015	0.073
	(0.042)	(0.041)	(0.049)	(0.052)	(0.051)	(0.048)	(0.043)	(0.052)	(0.042)	(0.051)
Duration				0.006	0.002					
				(0.007)	(0.009)					
Long, intense conflict					0.086					
					(0.141)					
Civilian fatalities (/1000)						0.283***	0.241***	0.241***		
						(0.072)	(0.072)	(0.072)		
Civilian fatalities, squared						-0.037***	-0.031***	-0.030***		
						(0.011)	(0.011)	(0.010)		
Battle-related fatalities (/1000)						0.029	0.028	0.027		
						(0.047)	(0.047)	(0.048)		
Battle-related fatalities, squared						-0.004	-0.004	-0.004		
						(0.003)	(0.003)	(0.003)		
Dummy: Sexual violence							0.207***	0.205***	0.211***	0.348***
							(0.066)	(0.070)	(0.065)	(0.097)
Freedom of religion			-0.070	-0.082	-0.087	-0.020	-0.035	-0.032	-0.078	-0.122
			(0.067)	(0.070)	(0.071)	(0.067)	(0.074)	(0.079)	(0.073)	(0.089)
Polity Indicator			-0.001	-0.001	-0.001	-0.002*	-0.002*	-0.002*	-0.001	-0.001
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Political instability			0.099	0.112	0.115	0.105	0.084	0.082	0.077	0.123
			(0.096)	(0.097)	(0.097)	(0.093)	(0.091)	(0.092)	(0.092)	(0.110)
GDP per capita, logged			-0.156	-0.169	-0.160	-0.169	-0.216	-0.215	-0.204	-0.297**
			(0.137)	(0.138)	(0.135)	(0.136)	(0.144)	(0.144)	(0.146)	(0.130)
Population size, logged			-0.372	-0.151	-0.040	-0.290	0.100	0.110	0.041	-0.257
			(0.775)	(0.801)	(0.900)	(0.781)	(0.753)	(0.751)	(0.731)	(1.187)
Child mortality			0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.001
			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Ln, internally displaced										0.017*
										(0.009)
Freedom from torture								-0.011		
								(0.053)		
Observations	9,186	9,186	9,186	9,186	9,186	9,186	9,186	9,186	9,186	6,664
Countrypairs	848	848	848	848	848	848	848	848	848	625
Years	14	14	14	14	14	14	14	14	14	14
Log likelihood	-3,681	-3,681	-3,681	-3,681	-3,681	-3,681	-3,681	-3,681	-3,681	-2,681
Pseudo-R2	0.124	0.125	0.126	0.126	0.126	0.126	0.126	0.126	0.127	0.129

Notes: Dependent variable: Share of female refugees. Fixed effects included are: Origin fixed effects, dyad fixed effects and destination-year fixed effects. Clustered (by countrypair and year) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 2: *Timing of the decision to leave and the composition of flows*

	(1)	(2)	(3)	(4)	(5)	(6)
Fatalities (/1000)	0.117*** (0.031)		0.087*** (0.034)	0.074** (0.036)	0.073* (0.039)	0.117*** (0.032)
Fatalities (/1000), squared	-0.007*** (0.002)		-0.006** (0.003)	-0.005* (0.003)	-0.005* (0.003)	-0.008*** (0.002)
Fatalities (/1000), t-1		0.145*** (0.033)	0.087*** (0.026)	0.101*** (0.029)	0.105*** (0.036)	
Fatalities (/1000),squared, t-1		-0.010*** (0.002)	-0.006*** (0.002)	-0.007*** (0.002)	-0.007*** (0.002)	
Fatalities (/1000), t-2			0.010 (0.022)	0.011 (0.022)	0.007 (0.023)	
Fatalities (/1000),squared, t-2			-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	
Freedom from political killings	0.015 (0.042)		0.004 (0.050)	0.009 (0.050)	0.016 (0.055)	0.015 (0.048)
Freedom of religion	-0.078 (0.073)		-0.080 (0.076)	-0.088 (0.072)	-0.111 (0.069)	-0.090 (0.073)
Polity Indicator	-0.001 (0.001)		-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.002** (0.001)
Political instability	0.077 (0.092)		0.051 (0.095)	-0.000 (0.107)	-0.021 (0.107)	0.024 (0.098)
GDP per capita, logged	-0.204 (0.146)		-0.167 (0.141)	0.368 (0.319)	0.331 (0.340)	-0.251* (0.138)
Population size, logged	0.041 (0.731)		0.190 (0.799)	0.201 (0.855)	0.440 (0.882)	0.221 (0.819)
Child mortality	0.003 (0.003)		0.003 (0.003)	0.004 (0.003)	0.004 (0.004)	0.003 (0.003)
Dummy: Sexual violence	0.211*** (0.065)		0.198*** (0.065)	0.201*** (0.072)	0.201*** (0.071)	0.226*** (0.068)
Polity Indicator, t-1		-0.002** (0.001)				
Political instability, t-1		0.038 (0.125)				
Freedom of religion, t-1		-0.069 (0.075)				
Freedom from political killings, t-1		-0.001 (0.057)				
Sexual violence, t-1		0.092* (0.053)			0.063 (0.053)	
GDP per capita, logged, t-1		-0.257** (0.110)		-0.603* (0.320)	-0.596* (0.333)	
Population, logged, t-1		0.051 (0.850)				
Mortality rate under 5, t-1		0.004 (0.003)				
Observations	9,186	8,600	9,186	9,038	8,600	8,600
Countrypairs	848	801	848	824	801	801
Years	14	14	14	14	14	14
Log likelihood	-3,681	-3,474	-3,681	-3,624	-3,474	-3,474
Pseudo-R2	0.127	0.126	0.127	0.126	0.127	0.126

Notes: Dependent variable: Share of female refugees. Fixed effects included are: Origin fixed effects, dyad fixed effects and destination-year fixed effects. Clustered (by countrypair and year) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 3: *Push factors: Where do compositional effects come from?*

	(1)	(2)	(3)
Fatalities (/1000)	0.813*** (0.161)	0.662*** (0.081)	
Fatalities (/1000), squared	-0.030*** (0.008)	-0.023** (0.010)	
Fatalities (/1000), t-1		0.373 (0.251)	
Fatalities (/1000),squared, t-1		-0.021 (0.017)	
Civilian fatalities (/1000)			3.095** (1.357)
Civilian fatalities, squared			-0.641** (0.275)
Battle-related fatalities (/1000)			0.595** (0.245)
Battle-related fatalities, squared			-0.014 (0.009)
Freedom from political killings	-0.081 (0.358)	-0.141 (0.354)	0.059 (0.385)
Freedom of religion	0.266 (0.434)	0.039 (0.358)	0.794 (0.688)
Polity Indicator	0.002 (0.006)	0.006 (0.007)	0.003 (0.007)
GDP per capita, logged	-0.439 (2.690)	0.100 (2.843)	-0.211 (1.831)
Population size, logged	-7.475 (7.728)	-9.158 (7.825)	-6.284 (8.256)
Child mortality	0.013 (0.017)	0.017 (0.018)	0.009 (0.015)
Dummy: Sexual violence	-0.023 (0.635)	-0.225 (0.678)	-0.036 (0.605)
Political instability	1.848*** (0.717)	1.987** (0.806)	1.881*** (0.725)
GDP per capita, logged, t-1	-0.524 (3.009)	-0.589 (3.128)	-0.655 (2.312)
Female=1	-0.686*** (0.227)	-0.644*** (0.248)	-0.625*** (0.214)
Female=1 × Fatalities (/1000)	0.000 (0.041)	-0.027 (0.048)	
Female=1 × Fatalities (/1000), squared	-0.001 (0.003)	-0.000 (0.003)	
Female=1 × Fatalities (/1000), t-1		0.038 (0.059)	
Female=1 × Fatalities (/1000),squared, t-1		-0.001 (0.004)	
Female=1 × Civilian fatalities (/1000)			-0.081 (0.154)
Female=1 × Civilian fatalities, squared			0.033 (0.040)
Female=1 × Battle-related fatalities (/1000)			-0.018 (0.042)
Female=1 × Battle-related fatalities, squared			-0.000 (0.003)
Female=1 × Dummy: Sexual violence	0.170 (0.162)	0.146 (0.151)	0.151 (0.144)
Female=1 × Freedom of religion	-0.021 (0.066)	-0.018 (0.065)	-0.020 (0.048)
Female=1 × Polity Indicator	0.001 (0.001)	0.001 (0.002)	0.001 (0.001)
Female=1 × Political instability	-0.028 (0.118)	-0.035 (0.122)	0.011 (0.123)
Female=1 × Freedom from political killings	0.009 (0.051)	0.004 (0.044)	0.007 (0.048)
Female=1 × GDP per capita, logged	-0.099 (0.608)	0.031 (0.621)	-0.171 (0.571)
Female=1 × GDP per capita, logged, t-1	0.177 (0.603)	0.047 (0.616)	0.241 (0.570)
Female=1 × Population size, logged	-0.027 (0.047)	-0.045 (0.061)	-0.020 (0.044)
Female=1 × Child mortality	0.002** (0.001)	0.002* (0.001)	0.002* (0.001)
Observations	18,134	18,134	18,134
Clusters	14	14	14
Log likelihood	-31,779,756	-31,779,756	-31,779,756
Pseudo-R2	0.800	0.805	0.809

Notes: Dependent variable: Dyadic flows in a stacked dataset. Fixed effects included are: Origin fixed effects, dyad fixed effects and destination-year fixed effects. Clustered (by countrypair and year) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 4: *Distances and composition of refugee flows*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance, logged, population weighted	-0.399*** (0.041)								
Distance(/1000)		-0.108*** (0.014)	-0.268*** (0.033)	-0.271*** (0.031)	-0.291*** (0.031)	-0.273*** (0.032)	-0.146*** (0.026)	-0.216*** (0.035)	-0.204*** (0.033)
Distance(/1000), squared			0.015*** (0.002)	0.015*** (0.002)	0.016*** (0.002)	0.015*** (0.002)	0.008*** (0.002)	0.012*** (0.003)	0.011*** (0.002)
Contiguity	0.121** (0.051)	0.314*** (0.047)	0.170*** (0.049)	0.167*** (0.051)	0.160*** (0.051)	0.116** (0.048)	-0.200*** (0.054)	0.018 (0.070)	-0.027 (0.058)
Colonial tie				-0.017 (0.087)	-0.006 (0.089)	0.003 (0.088)	0.087 (0.080)	0.004 (0.086)	-0.022 (0.086)
Shared official language				0.129** (0.062)	0.142** (0.059)	0.140** (0.058)	0.099* (0.055)	0.109* (0.063)	0.089 (0.059)
Religious distance				0.021 (0.064)	0.015 (0.066)	0.025 (0.067)	0.029 (0.063)	-0.000 (0.074)	0.003 (0.071)
Linguistic distance				0.079 (0.200)	0.004 (0.194)	0.018 (0.191)	0.071 (0.181)	0.124 (0.173)	0.077 (0.164)
Genetic distance, logged					0.070* (0.037)	0.080** (0.036)	0.066** (0.032)	0.078* (0.045)	0.074 (0.048)
Diaspora registered immigrants,logged, 2000						0.015* (0.009)	0.006 (0.009)	-0.003 (0.008)	-0.000 (0.008)
Diaspora, t-3							0.116*** (0.012)		
Female diaspora, average over 3 years								0.044*** (0.015)	
Male diaspora, average over 3 years									0.050*** (0.014)
Observations	17,792	17,792	17,792	17,792	17,792	17,792	17,792	10,673	11,246
Clusters	15	15	15	15	15	15	15	15	15
Log likelihood	-6,239	-6,239	-6,239	-6,239	-6,239	-6,239	-6,239	-4,435	-4,635
Pseudo-R2	0.084	0.083	0.085	0.085	0.085	0.086	0.093	0.078	0.079

Notes: Dependent variable: Share of female refugees. Fixed effects included are: Origin-year fixed effects, and destination-year fixed effects. Clustered (by country-pair and year) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Cost factors: Where do compositional effects come from?

	(1)	(2)	(3)
Distance(/1000)	-0.856*** (0.308)	-0.439 (0.562)	-0.571 (0.571)
Distance(/1000), squared	-0.031 (0.049)	-0.287 (0.203)	-0.240 (0.203)
Contiguity=1	2.772*** (0.333)	3.837*** (0.799)	3.793*** (0.785)
Colonial tie=1	0.201 (0.349)	-0.060 (0.348)	0.151 (0.398)
Shared official language=1	0.267 (0.325)	0.141 (0.474)	0.015 (0.446)
Religious distance	-0.969*** (0.278)	-1.346*** (0.419)	-1.347*** (0.389)
Linguistic distance	-2.317*** (0.636)	-2.071*** (0.578)	-2.565*** (0.668)
Genetic distance, logged	-0.517** (0.260)	-0.180 (0.377)	-0.176 (0.381)
Female diaspora, average over 3 years		-0.233** (0.106)	
Male diaspora, average over 3 years			-0.228** (0.114)
Female=1	-0.603 (0.419)	-0.490 (0.443)	-0.524 (0.414)
Female=1 × Distance(/1000)	0.118 (0.154)	0.088 (0.173)	0.092 (0.155)
Female=1 × Distance(/1000), squared	-0.058* (0.032)	-0.064* (0.037)	-0.061* (0.032)
Female=1 × Contiguity=0	0.000 (.)	0.000 (.)	0.000 (.)
Female=1 × Contiguity=1	0.225** (0.112)	0.391*** (0.112)	0.394*** (0.112)
Female=1 × Colonial tie=1	-0.484*** (0.141)	-0.556*** (0.129)	-0.555*** (0.120)
Female=1 × Shared official language=1	0.029 (0.085)	0.058 (0.090)	0.066 (0.083)
Female=1 × Religious distance	0.116 (0.122)	0.093 (0.113)	0.090 (0.120)
Female=1 × Linguistic distance	0.170 (0.118)	0.294** (0.115)	0.311*** (0.110)
Female=1 × Genetic distance, logged	-0.011 (0.043)	0.005 (0.049)	-0.003 (0.045)
Female=1 × Female diaspora, average over 3 years		-0.026** (0.011)	
Female=1 × Male diaspora, average over 3 years			-0.029*** (0.011)
Observations	35,584	21,510	22,558
Clusters	15	15	15
Log likelihood	-50,320,718	-41,485,841	-42,337,310
Pseudo-R2	0.747	0.750	0.748

Notes: Dependent variable: Dyadic flows in a stacked dataset. Fixed effects included are: Origin-year fixed effects, and destination-year fixed effects. Clustered (by country-pair and year) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 6: *Sorting and composition of refugee flows*

	(1)	(2)	(3)	(4)	(5)	(6)
Fatalities, logged, t-1	0.008 (0.010)	0.007 (0.011)	0.006 (0.011)	0.009 (0.011)	0.004 (0.012)	0.001 (0.013)
Peace duration, t-1	0.011* (0.006)	0.011* (0.006)	0.011* (0.006)	0.012* (0.006)	0.013** (0.006)	0.013** (0.006)
Freedom of Religion, t-1	-0.008 (0.099)	-0.015 (0.095)	-0.035 (0.095)	-0.038 (0.090)	-0.013 (0.097)	-0.006 (0.098)
Freedom of political killings	0.048 (0.061)	0.031 (0.070)	0.021 (0.069)	0.013 (0.067)	0.022 (0.065)	0.008 (0.065)
Polity, t-1		0.007 (0.015)	0.005 (0.015)	0.005 (0.015)	0.004 (0.014)	0.004 (0.014)
Political instability, t-1		-0.115 (0.195)	-0.134 (0.193)	-0.106 (0.193)	-0.087 (0.197)	-0.073 (0.198)
Exclusion by social group			-1.934** (0.973)	-1.731* (0.958)	-1.417 (0.892)	-1.559* (0.871)
Freedom of movement, women, t-1				0.315** (0.124)	0.311** (0.124)	0.337*** (0.127)
Property Rights, women, t-1				-0.127 (0.135)	-0.078 (0.133)	-0.089 (0.135)
GDP per capita, logged, t-1					0.109 (0.170)	0.121 (0.176)
Population size, logged, t-1					-0.288 (0.526)	-0.224 (0.552)
Mortality rate, under 5					0.008** (0.003)	0.008*** (0.003)
Diaspora						-0.065** (0.029)
Observations	13,216	13,216	13,216	13,216	13,216	13,216
Dyads	931	931	931	931	931	931
Years	15	15	15	15	15	15
Log likelihood	-4,791	-4,791	-4,791	-4,791	-4,791	-4,791
Pseudo-R2	0.135	0.135	0.136	0.136	0.137	0.138

Notes: Dependent variable: Share of female refugees. Fixed effects included are: Origin-year fixed effects, dyad fixed effects and destination fixed effects. Clustered (by country-pair and year) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 7: *Sorting, composition and (non-)neighbors*

	(1)	(2)	(Neighbor)	(Non-Neighbor)
Fatalities, logged, t-1	0.013 (0.016)	0.011 (0.017)	-0.041 (0.026)	0.009 (0.017)
Peace duration, t-1	0.017** (0.007)	0.017** (0.007)	0.001 (0.013)	0.017** (0.007)
Freedom of political killings	0.007 (0.085)	-0.006 (0.087)	-0.031 (0.125)	-0.007 (0.087)
Freedom of movement, women, t-1	0.301** (0.146)	0.321** (0.147)	0.566*** (0.185)	0.309** (0.146)
Property Rights, women, t-1	-0.104 (0.162)	-0.117 (0.162)	-0.229 (0.249)	-0.116 (0.163)
Exclusion by social group	-0.804 (0.999)	-0.965 (0.941)	-3.713** (1.465)	-0.754 (0.911)
Freedom of Religion, t-1	-0.005 (0.106)	-0.003 (0.109)	-0.081 (0.104)	0.010 (0.107)
Polity, t-1	0.009 (0.017)	0.010 (0.018)	0.002 (0.029)	0.009 (0.017)
Political instability, t-1=1	0.000 (0.231)	0.013 (0.228)	-0.347 (0.241)	0.013 (0.219)
GDP per capita, logged, t-1	0.195 (0.215)	0.207 (0.226)	0.075 (0.260)	0.180 (0.236)
Population size, logged, t-1	-0.088 (0.520)	-0.033 (0.543)	-2.412 (1.755)	-0.045 (0.540)
Mortality rate, under 5	0.008** (0.004)	0.009** (0.004)	0.006 (0.004)	0.009** (0.004)
Diaspora		-0.057 (0.042)	-0.073** (0.036)	-0.061 (0.043)
Contiguity=1 × Fatalities, logged, t-1	-0.042 (0.030)	-0.044 (0.031)		
Contiguity=1 × Peace duration, t-1	-0.028** (0.014)	-0.027* (0.014)		
Contiguity=1 × Freedom of political killings	0.062 (0.161)	0.061 (0.165)		
Contiguity=1 × Freedom of movement, women, t-1	0.030 (0.200)	0.041 (0.198)		
Contiguity=1 × Property Rights, women, t-1	0.099 (0.288)	0.117 (0.287)		
Contiguity=1 × Exclusion by social group	-2.935* (1.668)	-2.711* (1.623)		
Contiguity=1 × Freedom of Religion, t-1	-0.041 (0.102)	-0.024 (0.100)		
Contiguity=1 × Polity, t-1	-0.024 (0.029)	-0.022 (0.029)		
Contiguity=1 × Political instability, t-1=1	-0.456* (0.236)	-0.459* (0.235)		
Contiguity=1 × GDP per capita, logged, t-1	-0.283 (0.286)	-0.275 (0.295)		
Contiguity=1 × Population size, logged, t-1	-0.356 (1.245)	-0.308 (1.157)		
Contiguity=1 × Mortality rate, under 5	-0.003 (0.005)	-0.002 (0.005)		
Contiguity=1 × Diaspora		-0.020 (0.039)		
Observations	13,216	13,216	2,178	11,038
Dyads	931	931	152	779
Years	15	15	15	15
Log likelihood	-4,791	-4,791	-966	-3,802
Pseudo-R2	0.138	0.139	0.135	0.140

Notes: Dependent variable: Share of female refugees. Fixed effects included are: Origin-year fixed effects, dyad fixed effects and destination fixed effects. Clustered (by country-pair and year) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 8: *Sorting—where do compositional effects come from?*

	(All)	(Neighbors)	(Non-neighbors)
Female=1	0.949* (0.504)	-0.114 (0.421)	2.416*** (0.636)
Fatalities, logged, t-1	-0.189** (0.095)	-0.217** (0.105)	0.047 (0.100)
Peace duration, t-1	-0.082 (0.097)	-0.178 (0.117)	0.089*** (0.025)
Freedom of political killings	-0.551 (0.531)	-0.606 (0.575)	0.733* (0.431)
Exclusion	-4.302 (4.585)	-3.581 (5.178)	-10.382** (5.105)
Polity, t-1	-0.065 (0.097)	-0.095 (0.102)	-0.052 (0.091)
Political instability, t-1=1	-1.862** (0.783)	-2.094** (0.974)	-0.176 (0.825)
Freedom of movement, women, t-1	0.852 (1.377)	1.087 (1.425)	-0.762 (0.855)
Property Rights, women, t-1	-0.338 (0.796)	-0.715 (0.922)	2.943*** (1.112)
GDP per capita, logged, t-1	-1.230 (1.093)	-1.355 (1.244)	-0.104 (0.842)
Population size, logged, t-1	5.704 (10.984)	19.438 (13.482)	2.449 (4.304)
Mortality rate, under 5	-0.011 (0.016)	-0.015 (0.019)	-0.032 (0.022)
Diaspora	-0.215* (0.116)	-0.247** (0.117)	-0.225* (0.117)
Female=1 × Fatalities, logged, t-1	-0.020 (0.035)	-0.019 (0.037)	-0.032 (0.034)
Female=1 × Peace duration, t-1	-0.026 (0.016)	-0.013 (0.017)	-0.001 (0.009)
Female=1 × Freedom of political killings	-0.066 (0.060)	-0.067 (0.059)	-0.170 (0.119)
Female=1 × Exclusion	0.340 (0.308)	0.353 (0.307)	-0.946 (0.644)
Female=1 × Polity, t-1	0.014 (0.017)	0.021 (0.017)	-0.005 (0.013)
Female=1 × Political instability, t-1=1	-0.915 (0.880)	-0.979 (1.072)	-0.357** (0.148)
Female=1 × Freedom of movement, women, t-1	0.007 (0.079)	0.036 (0.079)	-0.102 (0.111)
Female=1 × Property Rights, women, t-1	0.064* (0.039)	0.030 (0.045)	0.182** (0.090)
Female=1 × GDP per capita, logged, t-1	-0.133** (0.055)	0.012 (0.051)	-0.298*** (0.091)
Female=1 × Population size, logged, t-1	-0.063 (0.047)	-0.055 (0.053)	-0.018 (0.057)
Female=1 × Mortality rate, under 5	-0.002 (0.002)	-0.000 (0.001)	-0.001 (0.002)
Female=1 × Diaspora	0.012 (0.013)	0.008 (0.011)	0.008 (0.026)
Observations	26,432	4,356	22,076
Dyads	931	152	779
Years	15	15	15
Log likelihood	-35,486,435	-22,098,150	-3,335,036
Pseudo-R2	0.750	0.652	0.792

Notes: Dependent variable: Dyadic flows in a stacked dataset. Fixed effects included are: Origin-year fixed effects, dyad fixed effects and destination fixed effects. Clustered (by country-pair and year) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A-1: *Push factors and composition of flows with diaspora*

	(1)	(2)	(3)	(4)	(5)	(6)
Fatalities (/1000)	0.126*** (0.033)	0.125*** (0.032)			0.117*** (0.031)	0.110*** (0.031)
Fatalities (/1000), squared	-0.008*** (0.002)	-0.008*** (0.002)			-0.007*** (0.002)	-0.007*** (0.002)
Freedom from political killings	-0.012 (0.049)	-0.019 (0.050)	-0.016 (0.048)	0.009 (0.044)	0.009 (0.044)	0.072 (0.050)
Dummy: Sexual violence				0.199*** (0.067)	0.203*** (0.066)	0.345*** (0.094)
Freedom of religion	-0.070 (0.067)	-0.071 (0.068)	-0.020 (0.069)	-0.035 (0.076)	-0.078 (0.074)	-0.122 (0.090)
Polity Indicator	-0.001 (0.001)	-0.001 (0.001)	-0.002* (0.001)	-0.002* (0.001)	-0.001* (0.001)	-0.001 (0.001)
Political instability	0.099 (0.096)	0.102 (0.096)	0.108 (0.092)	0.087 (0.090)	0.081 (0.092)	0.125 (0.109)
GDP per capita, logged	-0.156 (0.137)	-0.118 (0.130)	-0.131 (0.129)	-0.178 (0.138)	-0.166 (0.139)	-0.289** (0.141)
Population size, logged	-0.372 (0.775)	-0.273 (0.808)	-0.191 (0.807)	0.175 (0.765)	0.116 (0.750)	-0.226 (1.272)
Child mortality	0.003 (0.003)	0.003 (0.003)	0.003 (0.003)	0.002 (0.003)	0.002 (0.003)	0.001 (0.004)
Ln, internally displaced						0.017* (0.009)
Diaspora, t-3		0.060* (0.032)	0.060* (0.031)	0.056* (0.031)	0.056* (0.032)	0.009 (0.036)
Civilian fatalities (/1000)			0.288*** (0.072)	0.248*** (0.073)		
Civilian fatalities, squared			-0.039*** (0.011)	-0.032*** (0.011)		
Battle-related fatalities (/1000)			0.027 (0.046)	0.027 (0.047)		
Civilian fatalities, squared			-0.003 (0.003)	-0.004 (0.003)		
Observations	9,186	9,186	9,186	9,186	9,186	6,664
Countrypairs	848	848	848	848	848	625
Years	14	14	14	14	14	14
Log likelihood	-3,681	-3,681	-3,681	-3,681	-3,681	-2,681
Pseudo-R2	0.126	0.126	0.126	0.127	0.127	0.129

Notes: Dependent variable: Share of female refugees. Fixed effects included are: Origin fixed effects, dyad fixed effects and destination-year fixed effects. Clustered (by countrypair and year) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A-2: *Push factors and composition of (non-negative) flows*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Fatalities (/1000)	0.021*	0.107**	0.084**	0.077*	0.076*				0.074*	0.057
	(0.013)	(0.042)	(0.041)	(0.040)	(0.041)				(0.039)	(0.040)
Fatalities (/1000), squared		-0.007**	-0.005	-0.004	-0.004				-0.004	-0.004
		(0.003)	(0.003)	(0.003)	(0.003)				(0.003)	(0.003)
Freedom from political killings	0.046	0.072*	0.076*	0.092**	0.105**	0.084**	0.107***	0.063	0.101**	0.160***
	(0.038)	(0.043)	(0.044)	(0.045)	(0.044)	(0.042)	(0.040)	(0.057)	(0.041)	(0.044)
Duration				0.007	-0.004					
				(0.008)	(0.009)					
Dummy: Sexual violence							0.193***	0.211***	0.193***	0.395***
							(0.050)	(0.054)	(0.052)	(0.098)
Freedom of religion			-0.054	-0.068	-0.081	-0.014	-0.032	-0.059	-0.065	-0.139*
			(0.064)	(0.063)	(0.061)	(0.061)	(0.063)	(0.064)	(0.070)	(0.084)
Polity Indicator			-0.002*	-0.001*	-0.001	-0.002**	-0.002**	-0.002**	-0.001*	-0.001
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Political instability			0.101	0.116	0.123	0.098	0.071	0.081	0.075	0.107
			(0.088)	(0.090)	(0.086)	(0.080)	(0.076)	(0.079)	(0.082)	(0.088)
GDP per capita, logged			-0.161	-0.174	-0.150	-0.172	-0.210	-0.223	-0.198	-0.238
			(0.168)	(0.166)	(0.163)	(0.169)	(0.171)	(0.176)	(0.170)	(0.163)
Population size, logged			-1.313	-1.049	-0.756	-1.176	-0.794	-0.859	-0.922	-1.339
			(0.958)	(1.060)	(1.066)	(1.017)	(1.005)	(1.020)	(0.934)	(1.203)
Child mortality			0.001	0.001	0.001	0.002	0.001	0.002	0.001	-0.004
			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.005)
Ln, internally displaced										0.018**
										(0.007)
Long, intense conflict					0.226**					
					(0.108)					
Civilian fatalities (/1000)						0.228***	0.191**	0.191**		
						(0.082)	(0.076)	(0.077)		
Civilian fatalities, squared						-0.024	-0.018	-0.019		
						(0.015)	(0.014)	(0.014)		
Battle-related fatalities (/1000)						-0.014	-0.018	-0.012		
						(0.041)	(0.041)	(0.040)		
Civilian fatalities, squared						0.000	0.000	0.000		
						(0.003)	(0.003)	(0.003)		
Freedom from torture								0.089		
								(0.056)		
Observations	6,660	6,660	6,660	6,660	6,660	6,660	6,660	6,660	6,660	4,842
Countrypairs	810	810	810	810	810	810	810	810	810	595
Years	14	14	14	14	14	14	14	14	14	14
Log likelihood	-3,050	-3,050	-3,050	-3,050	-3,050	-3,050	-3,050	-3,050	-3,050	-2,229
Pseudo-R2	0.168	0.168	0.169	0.169	0.169	0.169	0.170	0.170	0.170	0.171

Notes: Negative flows dropped. Clustered (by countrypair) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A-3: *Timing of the decision to leave (Only non-negative flows)*

	(1)	(2)	(3)	(4)	(5)	(6)
Fatalities (/1000)	0.074* (0.039)		0.054 (0.039)	0.043 (0.041)	0.044 (0.044)	0.077* (0.040)
Fatalities (/1000), squared	-0.004 (0.003)		-0.003 (0.003)	-0.002 (0.003)	-0.002 (0.003)	-0.005 (0.003)
Fatalities (/1000), t-1		0.110*** (0.036)	0.066* (0.034)	0.080** (0.034)	0.077* (0.042)	
Fatalities (/1000),squared, t-1		-0.007*** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)	-0.006*** (0.003)	
Fatalities (/1000), t-2			0.024 (0.023)	0.026 (0.023)	0.022 (0.024)	
Fatalities (/1000),squared, t-2			-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	
Freedom from political killings	0.101** (0.041)		0.087* (0.046)	0.109** (0.048)	0.113** (0.054)	0.101** (0.045)
Freedom of religion	-0.065 (0.070)		-0.068 (0.073)	-0.088 (0.063)	-0.111* (0.061)	-0.082 (0.069)
Polity Indicator	-0.001* (0.001)		-0.001 (0.001)	-0.001 (0.001)	-0.002 (0.001)	-0.002* (0.001)
Political instability	0.075 (0.082)		0.042 (0.073)	-0.023 (0.081)	-0.065 (0.076)	-0.001 (0.084)
GDP per capita, logged	-0.198 (0.170)		-0.161 (0.159)	0.813*** (0.308)	0.748** (0.306)	-0.243 (0.163)
Population size, logged	-0.922 (0.934)		-0.645 (1.032)	-0.259 (0.989)	0.140 (1.006)	-0.388 (0.924)
Child mortality	0.001 (0.003)		0.001 (0.003)	0.003 (0.003)	0.003 (0.003)	0.001 (0.003)
Dummy: Sexual violence	0.193*** (0.052)		0.182*** (0.053)	0.201*** (0.056)	0.208*** (0.053)	0.225*** (0.052)
Polity Indicator, t-1		-0.002** (0.001)				
Political instability, t-1		0.004 (0.103)				
Freedom of religion, t-1		-0.045 (0.067)				
Freedom from political killings, t-1		0.077 (0.052)				
Sexual violence, t-1		0.103 (0.063)			0.066 (0.063)	
GDP per capita, logged, t-1		-0.329*** (0.098)		-1.051*** (0.268)	-1.032*** (0.258)	
Population, logged, t-1		-0.606 (0.898)				
Mortality rate under 5, t-1		0.002 (0.003)				
Observations	6,660	6,232	6,660	6,572	6,232	6,232
Countrypairs	810	774	810	797	774	774
Years	14	14	14	14	14	14
Log likelihood	-3,050	-2,877	-3,050	-3,012	-2,877	-2,877
Pseudo-R2	0.170	0.170	0.170	0.171	0.171	0.169

Notes: Negative flows dropped. Clustered (by countrypair) standard errors in parentheses. Significance levels: *** p<0.01 ** p<0.05 * p<0.1

Table A-4: *Push factors interacted with gender*

	(1)	(2)	(3)
Freedom from political killings	0.300 (0.311)	0.353 (0.295)	0.578** (0.262)
Freedom of religion	-0.059 (0.334)	-0.280 (0.352)	-0.240 (0.313)
Polity Indicator	0.004 (0.009)	0.006 (0.009)	0.002 (0.008)
GDP per capita, logged	-0.031 (1.179)	1.005 (1.378)	0.413 (0.971)
Population size, logged	-4.559 (6.377)	-5.110 (5.968)	-3.989 (6.115)
Child mortality	-0.022 (0.016)	-0.017 (0.014)	-0.024 (0.018)
Female=1	-0.749*** (0.224)	-0.715*** (0.241)	-0.654*** (0.223)
Fatalities (/1000)	0.580*** (0.090)	0.459*** (0.123)	
Fatalities (/1000), squared	-0.023*** (0.006)	-0.017** (0.008)	
Dummy: Sexual violence=1	-0.151 (0.373)	-0.079 (0.328)	0.163 (0.254)
Political instability=1	1.726*** (0.610)	1.727*** (0.618)	1.567*** (0.599)
GDP per capita, logged, t-1	-0.484 (1.347)	-1.117 (1.442)	-0.886 (1.267)
Female=1 × Fatalities (/1000)	0.003 (0.034)	-0.019 (0.033)	
Female=1 × Fatalities (/1000), squared	-0.001 (0.002)	-0.000 (0.002)	
Female=1 × Dummy: Sexual violence=1	0.125 (0.095)	0.117 (0.106)	0.080 (0.094)
Female=1 × Freedom of religion	-0.015 (0.050)	-0.015 (0.052)	-0.013 (0.041)
Female=1 × Polity Indicator	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)
Female=1 × Political instability=1	0.023 (0.105)	0.022 (0.119)	0.068 (0.103)
Female=1 × Freedom from political killings	-0.014 (0.041)	-0.011 (0.042)	-0.009 (0.036)
Female=1 × GDP per capita, logged	-0.283 (0.481)	-0.246 (0.482)	-0.368 (0.427)
Female=1 × GDP per capita, logged, t-1	0.361 (0.500)	0.324 (0.498)	0.435 (0.438)
Female=1 × Population size, logged	-0.016 (0.038)	-0.030 (0.054)	-0.009 (0.046)
Female=1 × Child mortality	0.002*** (0.001)	0.002** (0.001)	0.002** (0.001)
Fatalities (/1000), t-1		0.217 (0.162)	
Fatalities (/1000), squared, t-1		-0.008 (0.010)	
Female=1 × Fatalities (/1000), t-1		0.037 (0.054)	
Female=1 × Fatalities (/1000), squared, t-1		-0.002 (0.004)	
Civilian fatalities (/1000)			1.067*** (0.386)
Civilian fatalities, squared			-0.266*** (0.093)
Battle-related fatalities (/1000)			0.707*** (0.136)
Battle-related fatalities, squared			-0.026*** (0.006)
Female=1 × Civilian fatalities (/1000)			-0.026 (0.143)
Female=1 × Civilian fatalities, squared			0.021 (0.039)
Female=1 × Battle-related fatalities (/1000)			-0.019 (0.031)
Female=1 × Battle-related fatalities, squared			-0.000 (0.002)
Observations	20,342	20,342	20,342
Clusters	14	14	14
Log likelihood	-32,226,161	-32,226,161	-32,226,161
Pseudo-R2	0.942	0.944	0.945

Notes: Negative flows dropped. Clustered (by countrypair) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table A-5: *Distances and composition of refugee flows*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance, logged, population weighted	-0.092*** (0.008)							
Contiguity	0.071*** (0.020)	0.087*** (0.019)	0.082*** (0.019)	0.082*** (0.019)	0.072*** (0.018)	-0.005 (0.014)	-0.028 (0.021)	-0.021 (0.020)
Colonial tie			0.032 (0.037)	0.035 (0.037)	0.036 (0.038)	-0.001 (0.031)	-0.048 (0.039)	-0.053 (0.040)
Shared official language			0.026*** (0.008)	0.026*** (0.008)	0.023*** (0.008)	0.005 (0.006)	0.035*** (0.012)	0.026** (0.011)
Religious distance			-0.000 (0.012)	-0.000 (0.012)	0.002 (0.012)	0.000 (0.009)	-0.006 (0.016)	-0.002 (0.014)
Linguistic distance			-0.163*** (0.038)	-0.178*** (0.042)	-0.178*** (0.042)	-0.075** (0.032)	-0.040 (0.057)	-0.103* (0.051)
Genetic distance, logged				0.008 (0.007)	0.011 (0.007)	0.000 (0.006)	0.011 (0.014)	0.012 (0.012)
Distance(/1000)		-0.059*** (0.006)	-0.052*** (0.006)	-0.054*** (0.006)	-0.049*** (0.007)	-0.017*** (0.005)	-0.037*** (0.010)	-0.038*** (0.009)
Distance(/1000), squared		0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.001*** (0.000)	0.002** (0.001)	0.002*** (0.001)
Diaspora registered immigrants, logged, 2000					0.004** (0.002)	0.000 (0.001)	-0.001 (0.002)	0.001 (0.002)
Diaspora, t-3						0.039*** (0.002)		
Female diaspora, average over 3 years							0.042*** (0.003)	
Male diaspora, average over 3 years								0.041*** (0.003)
Observations	12,471	12,471	12,471	12,471	12,471	12,471	5,282	5,926
Clusters	15	15	15	15	15	15	15	15
Log likelihood	1,148	1,148	1,148	1,148	1,148	1,148	-473	-480
Pseudo-R2								

Notes: Negative flows dropped. Clustered (by countrypair) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A-6: *Distances and composition of refugee flows: Interact with neighbor*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Contiguity	0.082*** (0.019)	0.091*** (0.019)						
Colonial tie	0.035 (0.037)	0.069 (0.047)	0.038 (0.037)	0.035 (0.037)	0.034 (0.037)	0.036 (0.037)	0.039 (0.037)	0.036 (0.038)
Shared official language	0.026*** (0.008)	0.026*** (0.009)	0.027*** (0.008)	0.027*** (0.008)	0.026*** (0.008)	0.026*** (0.007)	0.022** (0.008)	0.023*** (0.008)
Religious distance	-0.000 (0.012)	-0.001 (0.013)	-0.002 (0.013)	-0.002 (0.013)	-0.001 (0.012)	-0.000 (0.012)	0.001 (0.012)	0.003 (0.012)
Linguistic distance	-0.178*** (0.042)	-0.215*** (0.043)	-0.180*** (0.041)	-0.177*** (0.042)	-0.175*** (0.042)	-0.174*** (0.042)	-0.179*** (0.041)	-0.291*** (0.090)
Genetic distance, logged	0.008 (0.007)	0.013 (0.008)	0.008 (0.007)	0.009 (0.007)	0.008 (0.008)	0.006 (0.007)	0.011 (0.007)	0.011 (0.007)
Distance(/1000)	-0.054*** (0.006)	-0.055*** (0.007)	-0.054*** (0.006)	-0.054*** (0.006)	-0.054*** (0.006)	-0.054*** (0.006)	-0.049*** (0.007)	-0.048*** (0.007)
Distance(/1000), squared	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.003*** (0.000)
Contiguity=1			0.026 (0.018)	0.038* (0.020)	0.023 (0.022)	0.026 (0.022)	0.136** (0.049)	0.076*** (0.018)
Long intense conflict,o=1			-0.000 (0.010)					
Contiguity=1 × Long intense conflict,o=1			0.087*** (0.022)					
Duration				0.000 (0.001)				
Contiguity=1 × Duration				0.004** (0.001)				
Insumfatalities					0.000 (0.001)			
Contiguity=1 × Insumfatalities					0.013*** (0.004)			
Dummy: Sexual violence						0.008 (0.005)		
Contiguity=1 × Dummy: Sexual violence						0.098*** (0.022)		
Diaspora registered immigrants, logged, 2000							0.005** (0.002)	-0.011 (0.010)
Contiguity=1 × Diaspora							-0.008 (0.005)	
Linguistic distance × Diaspora								0.016 (0.011)
Observations	12,471	11,222	12,471	12,471	12,471	12,471	12,471	12,471
Clusters	15	15	15	15	15	15	15	15
Log likelihood	1,148	997	1,148	1,148	1,148	1,148	1,148	1,148
Pseudo-R2								

Notes: Negative flows dropped. Clustered (by countrypair) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A-7: Cost factors interacted with gender, negatives missing

	(1)	(2)	(3)
Female=1	-1.128*** (0.238)	-1.036*** (0.262)	-1.116*** (0.274)
Distance(/1000)	-1.458*** (0.438)	-1.004** (0.450)	-0.748 (0.463)
Distance(/1000), squared	0.054 (0.046)	0.041 (0.036)	-0.032 (0.080)
Contiguity=1	2.939*** (0.562)	1.936** (0.840)	2.010** (0.894)
Colonial tie=1	-0.600 (0.736)	-0.354 (0.723)	0.172 (0.886)
Shared official language=1	0.120 (0.365)	0.088 (0.345)	-0.144 (0.405)
Religious distance	-0.698** (0.352)	-0.623** (0.309)	-0.721* (0.371)
Linguistic distance	-2.367*** (0.785)	-1.834*** (0.659)	-3.236*** (0.925)
Genetic distance, logged	-0.815*** (0.315)	-0.786*** (0.285)	-0.496 (0.417)
Female=1 × Distance(/1000)	0.381*** (0.127)	0.392*** (0.129)	0.362*** (0.130)
Female=1 × Distance(/1000), squared	-0.124*** (0.047)	-0.116*** (0.044)	-0.114** (0.045)
Female=1 × Contiguity=1	0.304** (0.141)	0.478*** (0.141)	0.505*** (0.146)
Female=1 × Colonial tie=1	-0.499*** (0.139)	-0.572*** (0.108)	-0.548*** (0.113)
Female=1 × Shared official language=1	0.157** (0.078)	0.121* (0.073)	0.169** (0.073)
Female=1 × Religious distance	0.179* (0.094)	0.167** (0.070)	0.185** (0.083)
Female=1 × Linguistic distance	0.335*** (0.058)	0.325*** (0.072)	0.473*** (0.064)
Female=1 × Genetic distance, logged	-0.029 (0.034)	-0.060** (0.027)	-0.017 (0.036)
Diaspora, t-3		0.218 (0.136)	
Female=1 × Diaspora, t-3		-0.034*** (0.013)	
Male diaspora, average over 3 years			0.171 (0.137)
Female=1 × Male diaspora, average over 3 years			-0.027 (0.017)
Observations	15,038	15,038	8,846
Clusters	15	15	15
Log likelihood	-29,076,642	-29,076,642	-24,568,172
Pseudo-R2	0.873	0.877	0.894

Notes: Clustered (by countrypair) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A-8: *Sort - robustness*

	(1)	(2)	(3)	(4)
Fatalities, logged, t-1	0.004 (0.012)	0.006 (0.014)	-0.001 (0.050)	-0.001 (0.050)
Peace duration, t-1	0.013** (0.006)	0.014** (0.007)	-0.001 (0.019)	-0.001 (0.019)
Freedom of Religion, t-1	-0.013 (0.097)	-0.007 (0.089)	-0.240* (0.130)	-0.240* (0.130)
Freedom of political killings	0.022 (0.065)	0.006 (0.073)	-0.156 (0.178)	-0.156 (0.178)
Polity, t-1	0.004 (0.014)	0.006 (0.015)	0.002 (0.034)	0.002 (0.034)
Political instability, t-1	-0.087 (0.197)	-0.060 (0.222)		
Exclusion by social group	-1.417 (0.892)	-1.745* (1.000)	-3.048** (1.518)	-3.048** (1.518)
Freedom of movement, women, t-1	0.311** (0.124)	0.187* (0.112)	0.851*** (0.156)	0.851*** (0.156)
Property Rights, women, t-1	-0.078 (0.133)	-0.011 (0.151)	-0.222 (0.315)	-0.222 (0.315)
GDP per capita, logged, t-1	0.109 (0.170)	0.086 (0.167)	-0.064 (0.373)	-0.064 (0.373)
Population size, logged, t-1	-0.288 (0.526)	-0.211 (0.523)	-1.995 (1.634)	-1.995 (1.634)
Mortality rate, under 5	0.008** (0.003)	0.006** (0.003)	-0.003 (0.005)	-0.003 (0.005)
Long and intense conflict			-0.024 (2.415)	-0.024 (2.415)
Political instability, t-1=1			0.044 (0.318)	0.044 (0.318)
Long and intense conflict × Fatalities, logged, t-1			-0.071 (0.052)	-0.071 (0.052)
Long and intense conflict × Peace duration, t-1			-0.003 (0.024)	-0.003 (0.024)
Long and intense conflict × Freedom of political killings			0.201 (0.190)	0.201 (0.190)
Long and intense conflict × Freedom of movement, women, t-1			-0.542*** (0.196)	-0.542*** (0.196)
Long and intense conflict × Property Rights, women, t-1			0.081 (0.236)	0.081 (0.236)
Long and intense conflict × Exclusion by social group			-1.228 (1.102)	-1.228 (1.102)
Long and intense conflict × Freedom of Religion, t-1			0.234* (0.139)	0.234* (0.139)
Long and intense conflict × Polity, t-1			0.000 (0.030)	0.000 (0.030)
Long and intense conflict × Political instability, t-1=1			-0.752* (0.388)	-0.752* (0.388)
Long and intense conflict × GDP per capita, logged, t-1			0.123 (0.319)	0.123 (0.319)
Long and intense conflict × Population size, logged, t-1			-0.356*** (0.078)	-0.356*** (0.078)
Long and intense conflict × Mortality rate, under 5			0.011* (0.006)	0.011* (0.006)
Observations	13,216	11,681	2,178	2,178
Dyads	931	825	152	152
Years	15	15	15	15
Log likelihood	-4,791	-4,217	-966	-966
Pseudo-R2	0.137	0.135	0.139	0.139

Notes: Clustered (by countrypair) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A-9: *Sort - fully interact with contig*

	(1)	(2)	(3)	(4)
Contiguity=1	0.000 (.)	0.000 (.)		
Fatalities, logged, t-1	0.008 (0.014)	0.009 (0.014)	-0.016 (0.018)	0.008 (0.014)
Peace duration, t-1	0.013* (0.008)	0.013* (0.008)	0.014 (0.009)	0.012* (0.007)
Freedom of political killings	-0.082 (0.085)	-0.076 (0.086)	-0.011 (0.119)	-0.075 (0.086)
Freedom of movement, women, t-1	0.381** (0.170)	0.367** (0.166)	0.218 (0.176)	0.358** (0.160)
Property Rights, women, t-1	-0.103 (0.205)	-0.094 (0.205)	0.078 (0.149)	-0.085 (0.206)
Exclusion by social group	-1.901** (0.949)	-1.812** (0.897)	-2.324* (1.403)	-1.604* (0.918)
Freedom of Religion, t-1	0.073 (0.096)	0.074 (0.095)	0.147 (0.137)	0.077 (0.096)
Polity, t-1	0.015 (0.017)	0.015 (0.017)	0.015 (0.023)	0.013 (0.017)
Political instability, t-1=1	0.234 (0.160)	0.231 (0.156)		
GDP per capita, logged, t-1	0.214 (0.215)	0.205 (0.216)	0.054 (0.198)	0.169 (0.234)
Population size, logged, t-1	-0.030 (0.542)	-0.059 (0.538)	-1.339 (1.576)	-0.134 (0.585)
Mortality rate, under 5	0.010*** (0.003)	0.010*** (0.004)	0.006 (0.004)	0.010*** (0.004)
Contiguity=1 × Fatalities, logged, t-1	-0.016 (0.023)	-0.016 (0.023)		
Contiguity=1 × Peace duration, t-1	-0.014 (0.014)	-0.014 (0.014)		
Contiguity=1 × Freedom of political killings	0.097 (0.117)	0.094 (0.117)		
Contiguity=1 × Freedom of movement, women, t-1	-0.255 (0.206)	-0.245 (0.201)		
Contiguity=1 × Property Rights, women, t-1	0.249 (0.244)	0.237 (0.246)		
Contiguity=1 × Exclusion by social group	-1.276 (1.587)	-1.351 (1.540)		
Contiguity=1 × Freedom of Religion, t-1	0.096 (0.142)	0.093 (0.140)		
Contiguity=1 × Polity, t-1	-0.005 (0.023)	-0.006 (0.023)		
Contiguity=1 × Political instability, t-1=1	-0.410** (0.190)	-0.407** (0.188)		
Political instability, t-1			-0.081 (0.223)	0.222 (0.156)
Contiguity=1 × GDP per capita, logged, t-1	-0.276 (0.235)	-0.259 (0.243)		
Contiguity=1 × Population size, logged, t-1	-0.500 (0.992)	-0.429 (1.008)		
Contiguity=1 × Mortality rate, under 5	-0.002 (0.005)	-0.002 (0.005)		
Diaspora		0.030 (0.037)	0.005 (0.033)	0.025 (0.039)
Contiguity=1 × Diaspora		-0.025 (0.035)		
Observations	9,998	9,998	1,401	8,597
Dyads	900	900	147	753
Years	15	15	15	15
Log likelihood	-4,078	-4,078	-760	-3,273
Pseudo-R2	0.188	0.188	0.186	0.183

Notes: Negative flows dropped. Clustered (by countrypair) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A-10: *Pull factors interacted with gender*

	(All)	(Neighbors)	(Non-neighbors)
Female=1	0.949*	-0.114	2.416***
	(0.504)	(0.421)	(0.636)
Fatalities, logged, t-1	-0.189**	-0.217**	0.047
	(0.095)	(0.105)	(0.100)
Peace duration, t-1	-0.082	-0.178	0.089***
	(0.097)	(0.117)	(0.025)
Freedom of political killings	-0.551	-0.606	0.733*
	(0.531)	(0.575)	(0.431)
Exclusion	-4.302	-3.581	-10.382**
	(4.585)	(5.178)	(5.105)
Polity, t-1	-0.065	-0.095	-0.052
	(0.097)	(0.102)	(0.091)
Political instability, t-1=1	-1.862**	-2.094**	-0.176
	(0.783)	(0.974)	(0.825)
Freedom of movement, women, t-1	0.852	1.087	-0.762
	(1.377)	(1.425)	(0.855)
Property Rights, women, t-1	-0.338	-0.715	2.943***
	(0.796)	(0.922)	(1.112)
GDP per capita, logged, t-1	-1.230	-1.355	-0.104
	(1.093)	(1.244)	(0.842)
Population size, logged, t-1	5.704	19.438	2.449
	(10.984)	(13.482)	(4.304)
Mortality rate, under 5	-0.011	-0.015	-0.032
	(0.016)	(0.019)	(0.022)
Diaspora	-0.215*	-0.247**	-0.225*
	(0.116)	(0.117)	(0.117)
Female=1 × Fatalities, logged, t-1	-0.020	-0.019	-0.032
	(0.035)	(0.037)	(0.034)
Female=1 × Peace duration, t-1	-0.026	-0.013	-0.001
	(0.016)	(0.017)	(0.009)
Female=1 × Freedom of political killings	-0.066	-0.067	-0.170
	(0.060)	(0.059)	(0.119)
Female=1 × Exclusion	0.340	0.353	-0.946
	(0.308)	(0.307)	(0.644)
Female=1 × Polity, t-1	0.014	0.021	-0.005
	(0.017)	(0.017)	(0.013)
Female=1 × Political instability, t-1=1	-0.915	-0.979	-0.357**
	(0.880)	(1.072)	(0.148)
Female=1 × Freedom of movement, women, t-1	0.007	0.036	-0.102
	(0.079)	(0.079)	(0.111)
Female=1 × Property Rights, women, t-1	0.064*	0.030	0.182**
	(0.039)	(0.045)	(0.090)
Female=1 × GDP per capita, logged, t-1	-0.133**	0.012	-0.298***
	(0.055)	(0.051)	(0.091)
Female=1 × Population size, logged, t-1	-0.063	-0.055	-0.018
	(0.047)	(0.053)	(0.057)
Female=1 × Mortality rate, under 5	-0.002	-0.000	-0.001
	(0.002)	(0.001)	(0.002)
Female=1 × Diaspora	0.012	0.008	0.008
	(0.013)	(0.011)	(0.026)
Observations	26,432	4,356	22,076
Dyads	931	152	779
Years	15	15	15
Log likelihood	-35,486,435	-22,098,150	-3,335,036

Notes: Clustered (by countrypair) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A-11: Pull factors interacted with gender - excluding transit countries

	(All)	(Neighbors)	(Non-neighbors)
Female=1	0.965* (0.505)	-0.148 (0.425)	2.585*** (0.670)
Fatalities, logged, t-1	-0.186* (0.096)	-0.215** (0.105)	0.059 (0.104)
Peace duration, t-1	-0.083 (0.097)	-0.178 (0.114)	0.094*** (0.025)
Freedom of political killings	-0.549 (0.541)	-0.604 (0.586)	0.774* (0.446)
Exclusion	-3.709 (4.645)	-2.908 (5.111)	-10.545** (5.282)
Polity, t-1	-0.068 (0.097)	-0.102 (0.102)	-0.056 (0.101)
Political instability, t-1=1	-1.959** (0.809)	-2.213** (1.055)	-0.190 (0.891)
Freedom of movement, women, t-1	0.863 (1.390)	1.097 (1.425)	-0.678 (0.934)
Property Rights, women, t-1	-0.236 (0.795)	-0.608 (0.905)	3.318*** (1.260)
GDP per capita, logged, t-1	-1.222 (1.112)	-1.345 (1.267)	-0.140 (0.843)
Population size, logged, t-1	7.154 (11.636)	21.992 (14.095)	3.177 (4.251)
Mortality rate, under 5	-0.010 (0.016)	-0.015 (0.018)	-0.035 (0.022)
Diaspora	-0.213* (0.116)	-0.243** (0.116)	-0.213* (0.121)
Female=1 × Fatalities, logged, t-1	-0.022 (0.036)	-0.020 (0.036)	-0.037 (0.036)
Female=1 × Peace duration, t-1	-0.025 (0.016)	-0.013 (0.017)	0.001 (0.009)
Female=1 × Freedom of political killings	-0.067 (0.060)	-0.066 (0.060)	-0.170 (0.125)
Female=1 × Exclusion	0.334 (0.314)	0.350 (0.319)	-0.993 (0.678)
Female=1 × Polity, t-1	0.014 (0.018)	0.022 (0.017)	-0.008 (0.015)
Female=1 × Political instability, t-1=1	-1.074 (0.980)	-1.202 (1.196)	-0.361** (0.149)
Female=1 × Freedom of movement, women, t-1	0.005 (0.081)	0.034 (0.079)	-0.116 (0.119)
Female=1 × Property Rights, women, t-1	0.063 (0.040)	0.030 (0.045)	0.199** (0.095)
Female=1 × GDP per capita, logged, t-1	-0.135** (0.055)	0.017 (0.052)	-0.312*** (0.093)
Female=1 × Population size, logged, t-1	-0.063 (0.047)	-0.053 (0.053)	-0.016 (0.056)
Female=1 × Mortality rate, under 5	-0.002 (0.002)	-0.000 (0.001)	-0.001 (0.002)
Female=1 × Diaspora	0.011 (0.014)	0.007 (0.011)	0.004 (0.025)
Observations	23,362	4,008	19,354
Dyads	825	140	685
Years	15	15	15
Log likelihood	-34,562,563	-21,587,713	-3,237,111
Pseudo-R2	0.747	0.649	0.793

Notes: Clustered (by countrypair) standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.