On the push and pull factors of internal and external forced migration

A thesis submitted to the University of the Basque Country
for the degree of Doctor of Philosophy in the
Faculty of Economics
2020

Jon Echevarria
Supervisor: Javier Gardeazabal
University of the Basque Country UPV/EHU
Fundamentos del Análisis Económico II

Contents

List of Figures

List of Tables

Part I

Acknowledgements

This thesis would not have been written without the emotional support and academic advice of many people. First of all, I would like to thank my supervisor, Javier Gardeazabal. It has been a considerable pleasure working with you all these years, to listen to your advice, ideas, corrections and instructions in order to complete this research. I very much appreciate the fact that you always made time in your busy schedule to meet up, to comment on every version of these chapters, or to answer emails. Besides, your support has been very important for me throughout the process of writing this thesis.

I would like to thank the Basque Government for funding this project and the University of the Basque Country for providing research facilities and an inspiring academic environment. I would also like to thank Daniel Arce, María Paz Espinosa, Khusrav Gaibulloev, Martin Gassebner, Eric Neumayer, Irene Olloqui, Clint Peinhardt, James Piazza, Todd Sandler, Fernando Sanz, William Shughart and Thomas Zeitzoff for very helpful comments.

The chapters in this thesis were presented at the 2016 and 2018 University of Texas at Dallas Political Violence and Policy Conference, the 2018 Australasian Meeting of the Econometric Society at Auckland University of Technology, the 2018 Meeting of the Spanish Economic Association (SAEe 2018) held at Getafe Campus at the Universidad Carlos III de Madrid, and in seminars at the Universidad de la República, Montevideo, the Refugee Studies Centre at the University of Oxford, the Economics faculty of Università Cattolica del Sacro Cuore in Milan and at the University of the Basque Country.

I would not have had the same energy and happiness writing this thesis without my family and friends. First, I thank my parents, Juan Luis and María José for making such an important investment in my academic education from my childhood right up to the current day, for teaching me values which have been key in my life, and for giving me your unconditional love. I also thank my sister, Leire, and my aunt, Mari, for always helping me to disconnect and for always trying to make me happy.

I have had some doubts and hard times through all these years, but my friends have given me all the energy needed to make these doubts disappear. Without them, this would not have been possible. Thank you, Ander. Our shared passion, football, has taught us that without the strength and power of others, no one can reach their goal.Throughout these years, you have showed me that together we can overcome any obstacle in our lives. Thank you Aitor. We have always been together, going to the same schools, living through so many moments together, and helping each other in every moment. Thank you for spending many hours with me while I was working on this research project, and for visiting me in my stays abroad.Joana, thank you for constantly encouraging me, for supporting me when I thought I was not able to continue with this research, and for showing me the importance of feminism and gender in our society. Part of this thesis is also yours. Thank you, Yon and Javi, among others, for always being concerned about my welfare, for asking me how my thesis was going and for making me feel special.

Without these people, all the academic process, from my school in Ermua right up to University, would have been much more difficult, and I would not have been so happy as I have been.

Part II

Motivation

Forced migration has been in the news over the last decade due to the remarkable rise in both refugees and internally displaced persons (IDPs). In order to analyze the evolution of forced migration, it is very important to determine which factors are the cause of internal or external forced migration, why forced migrants leave, and from which origin countries and to which destination countries they go. Millions of refugees and IDPs move away from their home locations every year in order to survive, and this number has increased exponentially over the last decades. Therefore, we crucially need to know the factors behind these movements to fight forced migration more effectively in the future.

This thesis studies the evolution of forced migration from three perspectives, which correspond to the three main chapters: (i) an analysis of dyadic refugee flows and their determinants, (ii) a spatial model of internal displacement in origin countries and (ii) the gender balance of refugee flows in destination countries.

The first chapter investigates the determinants of global forced migration using a gravity model, a workbench used to model international trade and migration. Most studies about forced migration are focused on a specific region, most often Western European Countries, and only based on destination flows. Therefore, it is difficult to get a clear explanation regarding refugee flows if only certain destination countries are considered, without taking into account from which origin countries forced migrants emigrate. In order to understand forced migration, it is essential to relate origin and destination countries, along with the characteristics of the origin country. In other words, the goal of this chapter is to determine the causes that lead to refugees leaving origin countries, and the factors which attract those refugees to specific destination countries. The motivation of this analysis also consists on the UNHCR data showing that the pattern of forced migration flows has changed. Over the last three decades, the evolution of refugee flows has grown exponentially, thus becoming a very relevant issue in our society. Thus, our goal is to model global patterns of refugee migration. Previous literature on forced migration has proven that factors like armed conflicts and the standards of democracy or civil liberties play a central role in the determination of forced migration, along with dyadic variables like distance between origin and destination countries, and sharing a common border or language.

The evidence reported below indicates that conflict in the source country, civil liberties in the origin and destination countries, population at destination countries, as well as proximity of source and destination country, are the most significant driving factors of forced migration. These results have one important policy implication: the only channel through which politicians can influence refugee counts is by external policy directed towards reducing conflict and improving the standards of civil liberties in the source countries.

The second chapter develops a spatial model of internal and external forced migration, reminiscent of Hotelling's spatial model in economics and Schelling's model of segregation. We study how refugees and IDPs are generated, that is, why some become refugees while other displaced persons remain in their countries of origin as IDPs. Forced displacement may have its origin in armed conflict, lack of civil liberties or political rights and natural disasters. However, internal and external displacement share a common cause: armed conflict. When we focus on displaced people who escape from armed conflict, the distinguishing feature between refugees and IDPs is whether they cross an international border. The evolution of both groups has been completely different over the last few decades: while the aggregate number of refugees increased steadily up to the beginning of the nineties, the aggregate number of IDPs has remarkably grown since 2000, surpassing the number of refugees, and currently standing at 41 million. As refugees and IDPs are driven by a common cause, it seems logical to study both groups of people simultaneously. However, the literature on forced displacement is mainly divided into refugee studies on the one hand,and internal displacement studies on the other. While refugees and IDPs share a common cause, their evolution over the last few decades is very different. Thus, the goal of this chapter is to investigate the role of spatial factors in determining the split into the two groups.

Furthermore, the model designed below builds a bridge between internal and external forced migration. As far as we know, simultaneous modeling of internal and external forced migration has not been attempted before.

The spatial model we develop represents armed conflict as a shock that takes place at a particular location and generates a migration flow. The model predicts how the number of refugees and IDPs, as a fraction of a country's population, varies with the intensity of armed conflict and geographical covariates such as country size, orography and distance to other countries. In our model, armed conflict has a causal effect on displacement, and the geographical variables exert a modification effect on displacement.

The final chapter studies the difference in the number of male and female refugees, i.e. the gender bias in destination flows. This chapter contributes a gender perspective to the forced migration literature. Most studies related to refugee destination flows do not distinguish between male and female refugees. UNHCR data shows that, for many destination countries, the number of male refugees is far greater than the number of female refugees. Therefore, it is very important to understand why male refugees tend to go in larger numbers than female refugees to certain destination countries. The situation of female refugees has been studied from many perspectives, but there are not studies relating to the gender profile of refugee destination flows, neither on a country-level, nor on a global level. This chapter is an attempt to fill this gap in the literature.

We study the extent to which refugee arrivals in destination countries exhibit a gender bias, and its determinants. In order to analyze the gender bias, we take into account factors such as civil liberties and destination country per capita GDP, country of origin-related variables weighted by distance, along with measures of women's rights.

Women's economic and social rights are worse than men's in most countries. This suggests that women will tend to find it more difficult to flee to certain countries with lower standards of women's rights. However, the evidence indicates that female refugees flee to countries where they have even fewer rights than in their origin countries.

Part III

The chapters

Chapter 1

Refugee generation

Abstract

This chapter makes use of a gravity model to investigate the determinants of global forced migration. We find that omitting zero counts results in parameter estimates that underestimate the effect of covariates on refugee counts. We compare the pooled regression, which does not account for unobserved heterogeneity, and the fixed-effects method, which does not identify the effect of time-invariant covariates, and cannot shed much light when covariates vary little. We propose the Pre-sample Mean Generalized Method of Moments (PSM-GMM) estimator, which accommodates the zeros, accounts for unobserved heterogeneity, but does not have the drawbacks of the fixed-effects methods when covariates exhibit little variation. In addition, using recently developed methods to estimate standard errors that are robust to dyadic correlation, we find evidence suggesting that previous findings based on underestimated standard errors mostly remain true after properly adjusting standard errors.In particular, conflict and civil liberties at the source country and proximity are found to be significant determinants of forced migration. However, some covariates previously found to be significant, such as sharing a common language or having a colonial relationship, lose significance. In addition, we find a significant positive influence of the level of civil liberties in the destination country on

the number of refugees it receives, a mechanism not explored before.

1.1 Introduction

Over the years, the United Nations High Commissioner for Refugees (UNHCR) has compiled a database on the number of refugee and refugee-like people, classified according to their country of origin and asylum/residence.[1]

This chapter explores these data by means of a gravity model of global forced migration. Contrary to the norm in forced migration studies where the analysis is typically restricted to a region, most often Western European Countries, this chapter studies global forced migration counts.

This analysis is motivated because the UNHCR data shows that the pattern of forced migration flows has changed. During the first decades after its inception in 1950, the number of origin and destination countries was small. However, nowadays almost all countries are the origin of refugees and most of them are also the destination of some refugees. Thus, our goal is to model global patterns of refugee migration.

The gravity model, a workbench used to model international trade and migration, stands as a natural way to describe this type of data. The specification of the covariates used in the gravity model borrows from previous research on forced migration.

Previous evidence has it that the existence of armed conflict and the standards of democracy or civil liberties play a central role in the determination of forced migration. In addition, researchers have found that dyadic magnitudes such as distance between countries and sharing a common border or language are fundamental determinants of forced migration.

A second distinguishing feature of this chapter is the type of data analyzed. While most of the empirical analysis in the literature focuses on asylum application data, our analysis uses

[1]The term asylum/residence is used by the UNHCR to account for those refugees to whom the asylum statute has been granted and also those to whom the asylum statute has been denied or simply did not apply for it. In the sequel, we will use the term "destination countries" to refer to asylum/residence countries.

UNHCR refugee counts, which include asylum applicants and also those people who are in a refugee-like situation but, for whatever the reason, did not apply for asylum.

These data are, therefore, less exposed to the problem often claimed for application data, i.e., the supposition that many asylum seekers are in fact economic migrants. This benefit comes at the cost of having less accurate data, as the UNHCR compiles them using not only administrative records from national governments, but also from Nongovernmental Organizations (NGOs) and UNHCR field officers estimates.

A third contribution of this chapter is methodological. Analysis of dyadic data, such as the refugee counts analyzed in this chapter, involves some issues on which not all researchers agree. We contribute to two of these debates on the analysis of dyadic data.

First, we look at the consequences of treating as missing or zeros all unobserved dyads. Second, we study the consequences of pooling together all observations versus allowing for dyad-specific unobserved effects, and propose a third route to account for unobserved heterogeneity using the Pre-sample Mean Generalized Method of Moments (PSM-GMM) estimation method.

A fourth contribution of this chapter is to assess whether previous empirical evidence on forced migration remains valid in light of the recently developed methods to make standard errors robust to inter-dyad correlation. For instance, in a directed dyadic relation, two observations might be correlated if they share common origin or destination.

Typically used standard errors that are clustered at the dyad are robust to intra-dyad correlation, but ignore inter-dyad correlation and overestimate the precision of the estimates. Montecarlo-based evidence, e.g. Cameron and Miller (2014), suggests that properly adjusted standard errors that are robust to inter-dyadic correlation could be several times larger than those that account for intra-dyadic correlation.

The evidence reported below indicates that conflict at the source country, civil liberties at the origin and destination countries, population at destination countries and proximity are the most significant driving factors of forced migration.

These results have one important policy implication: the only channel through which politicians can influence refugee counts is by external policy directed towards reducing conflict and improving the standards of civil liberties in the source countries. Although most of the driving forces were already known in the literature, we find evidence that as civil liberties improve at the destination countries, more refugees will choose that destination.

Section 2 provides a few references to the literature on forced migration. Section 3 describes the United Nations data on refugees. Section 4 describes the gravity model and Section 5 the methods. The main results are reported in Section 6, which is followed by some robustness checks in Section 7. Finally, Section 8 highlights the main conclusions and policy implications of this research.

1.2 Literature review

Theoretical studies on forced migration range from cost-sharing models, where the aim is to implement a way to share the burden among hosting countries, to behavioral models where refugees are subject to strict restrictions, but there is still some room for choice, e.g. Czaika (2009b). The burden share models, e.g., Czaika (2009a), and the empirical evidence, e.g., Roper and Barria (2010), suggest that states regard refugee protection by the UNHCR as an impure public good, as defined by Sandler and Hartley (2001). Namely, such protection provides both country-specific and purely public benefits.

Several previous empirical studies on forced migration have focused on asylum seeker applications in Western Europe. These include the analysis of asylum recognition rates, e.g., Holzer et al. (2000) and Neumayer (2005b), refugee origin, e.g., Neumayer (2005a) and destination choice, e.g., Havinga and Böcker (1999) and Neumayer (2004).

There are also empirical studies for African countries, e.g., Iqbal (2007), a continent that has witnessed large flows of interstate forced migration. Moore and Shellman (2007b) study destination patterns using global data as done here.

Scholars have not only studied the determinants of forced migration but also analyzed its effects. Inflows of forced migrants have been found to increase the likelihood of civil war in the asylum country, e.g., Salehyan and Gleditsch (2006). Refugee flows between states significantly increase the likelihood of militarized interstate conflict (e.g., Salehyan (2008)). Origin countries have been found to be the source country of transnational terrorism (e.g., Milton et al. (2013)).

On the contrary, there is evidence suggesting that the higher the inflow of refugees a country receives, the more likely it is to contribute to peacekeeping missions in the refugee source country (e.g., Uzonyi (2015)), and to contribute more Official Development Assistance to the refugee source country (e.g., Czaika (2009b)).

All the available empirical evidence indicates that armed conflict in the source country is the most important determinant of forced migration. Yet, perhaps the single clearest piece of evidence is Ball et al. (2002), who provide two-day frequency estimates of the number of killings that occurred in Kosovo from late March to mid-June 1999 and the number of refugees who migrated. Their plots show how these two magnitudes exhibit astonishing parallel trends.

A similarly important determinant of forced migration is the absence of civil liberties or political rights in the country of origin, (e.g., Neumayer (2004), Neumayer (2005b) and Moore and Shellman (2007b)). Previous research has found geographic proximity as well as sharing a common language and being a former colonizer to be determinants of forced migration, (e.g., Barthel and Neumayer (2015)).

1.3 Persons of concern data

According to the 1951 Refugee Convention and the 1967 Protocol, UNHCR (2011), a refugee is someone who "owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality, and is unable to, or owing to such fear, is unwilling to avail himself of the protection of that country."

The 1969 Organization of African Unity Convention governing the specific aspects of refugee problems in Africa, UNHCR (2006), accepted the definition of the 1951 Refugee Convention with the 1967 Protocol and extended it to include people who left their countries of origin not only because of persecution, but also due to acts of external aggression, occupation, domination by foreign powers or serious disturbances of public order.

The Convention does not apply to those who are believed to have committed war crimes, crimes against humanity, serious non-political crimes or those who benefit from protection from or assistance of a United Nations agency other than UNHCR, notably the United Nations Relief and Works Agency for Palestinian Refugees in the Near East, UNRWA. In addition to refugees, the UNHCR currently assists millions of other persons who do not fulfill the definition given above but are in a refugee-like situation.

In summary, a person can ask for refugee status for only two reasons: conflict or persecution in the country of origin. Therefore, the existence of conflict and the absence of civil liberties or political rights are the fundamental determinants of why a person would flee his or her country of origin to become a refugee.

The UNHCR provides two sources of data: the persons of concern data and the asylum seekers (applications) dataset. Both data sets are worth exploring. The asylum application data has been extensively analyzed elsewhere.[2]

We focus our analysis on the persons of concern data base. While the asylum application data is compiled by governmental agencies, the persons of concern database draws from information from those same governmental agencies and also from UNHCR field officers and NGOs.

In addition to having received less attention in the past, there are other reasons why we focus on the persons of concern database. First, the asylum application data might be affected by a sample selection problem. As argued by Jones (2009), refugee status determination varies

[2]See Neumayer (2004, 2005a, 2005b) and Barthel and Neumayer (2015).

considerably both between countries and also within country.

Knowing this, forced migrants might decide to self select into the application process, thus introducing a sample selection problem. In addition, it is often argued that some of the applicants might be economic migrants.

Second, the persons of concern database, however, includes data from asylum applications as well as from UNHCR field officers and NGOs' estimates on those cases where public authorities cannot provide the numbers.

The UNHCR's population of concern database provides counts of refugees; that is, those individuals recognized under the 1951 Convention, its 1967 Protocol, the 1969 OAU Convention as well as people in a refugee-like situation. It also includes Palestinian refugees living outside the UNRWA areas of operation. The UNHCR's database includes dyadic refugee counts for pairs of country of origin, asylum/residence country and year.[3]

The persons of concern database also includes other categories: Internally Displaced Persons (IDPs), asylum seekers (pending applications), stateless persons, returnees, and other persons of concern and resettlements. In the sequel, we will focus on UNHCR's data on refugees.

In some cases, UNHCR's data run from as early as 1951 to 2014, with the 2014 data being incomplete. We first look at the time trends of the aggregate data to give a broad picture of the data set. We add up counts of refugees across country of origin and residence to obtain a single annual time series.

As shown in Figure 1.1, the total number of refugees remained more or less stable during the 50s, 60s and early 70s when it started to rise steadily to 17.8 million in 1991. Thereafter, refugee counts started to fall.

[3]Crisp (1999) explain these data.

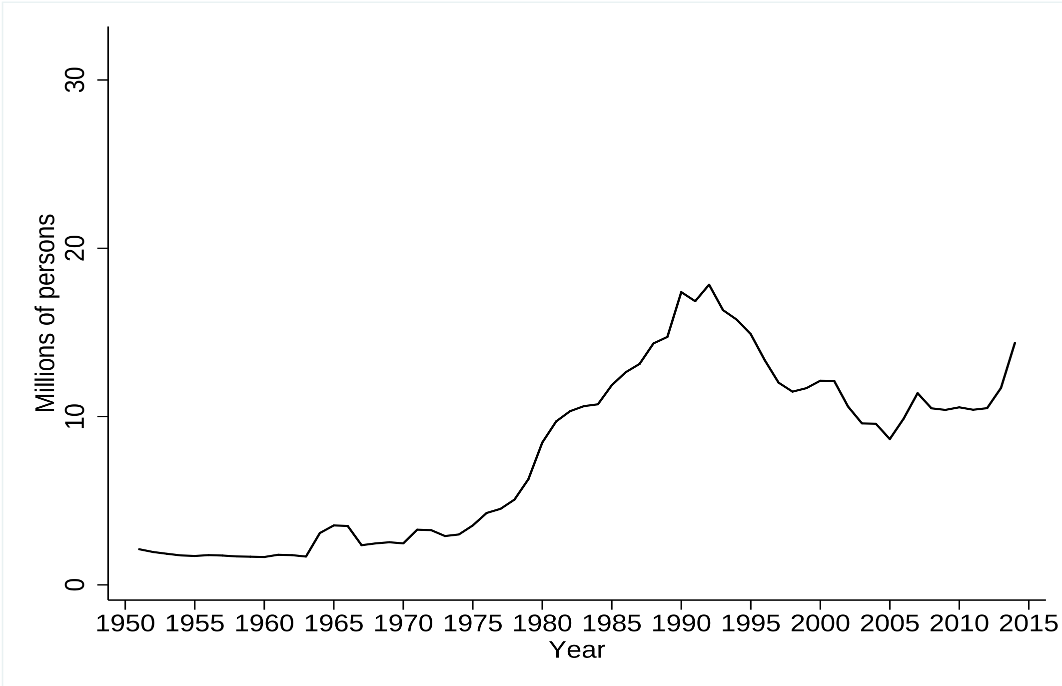

Figure 1.1: Trends in the aggregate number of refugees

Summary statistics reported in Table 1.1 indicate that there are 84,393 counts of refugees ranging from 1 to 3.272 million persons. A distinct mark of these data is that, refugee counts take a minimum value of one person, so there are no zeros in the data.

Table 1.1: Summary statistics

Variable	# obs.	mean	stand. dev.	min	max
Refugees	84,393	5,992	65,752	1	3,272,290

The absence of zeros may be due to the way the UNHCR collects the data. Governmental agencies, UNHCR field officers and NGOs report counts of persons to the UNHCR statistical office. Of course, these are positive counts.

Therefore, it is our understanding that those censored dyad counts are in fact zeros. The UNHCR started recording refugee counts in its first mandate in December 1950. The records for the early years contain only a few data points corresponding to the number of refugees in each country of asylum/residence while the origin country is unknown.

The first registered entry with a specific origin country is in 1960. Figure 1.2 plots the number of (known) countries of origin and asylum/residence per year. The number of countries of asylum/residence was above the number of countries of origin until 1990 when the pattern reversed.

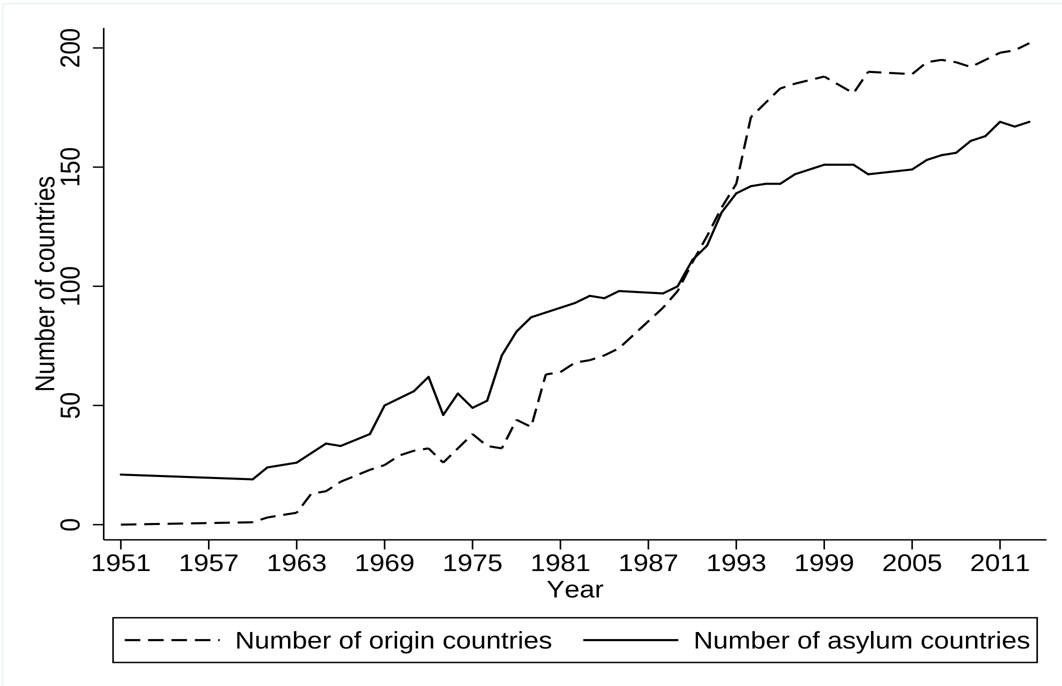

Figure 1.2: Number of countries of origin and residence by year

This graph shows how forced migration has experienced a tremendous change. What used to be a flow of people from conflicted, mostly developing countries now also originates from developed countries that are not experiencing armed conflict.

Figure 1.3 plots the number of refugees whose origin is the United States, United Kingdom and Sweden to all destination countries.[4] This figure shows that well-developed democratic countries are also the source of refugees.

Figure 1.4 shows the number of positive dyad counts per year. The annual number of dyad observations up until 1978 was always below 100, reached 1,000 by 1993 and almost 5,000 in 2013. If we simply multiply the number of countries of origin by the number of destination

[4]USA data exclude those refugees who flee to Canada, which are abnormally high due to the Safe Third Country Agreement between Canada and the United States. Under the Agreement, refugee claimants in each country are required to request refugee protection in the first safe country they arrive in, despite whether their final destination is another country.

countries, we obtain the potential number of dyad observations, which is also plotted in Figure 1.4.[5]

In 2013, we observe approximately 5,000 out of the 35,000 potential dyad observations, hence about 6 out of every 7 potential dyadic observations have been missed. Some of these "missing" observations could actually be missing observations, but as argued above, the bulk of them correspond to dyads where there are no flows of forced migrants. Therefore, we will treat them as zeros. In addition to refugee data, this chapter also uses data on other covariates whose sources and definitions are described in Appendix A.

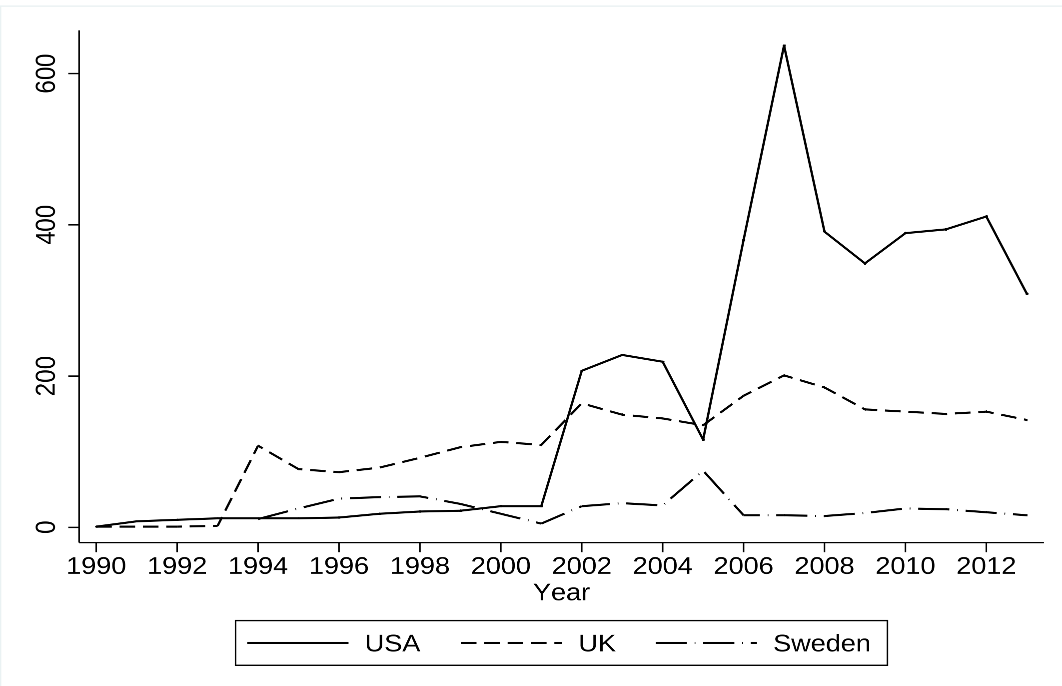

Figure 1.3: Number of refugees to all destinations from selected countries

[5]Notice that the theoretically feasible number of dyads is the square of the number of sovereign states, which is a larger number than what we call potential number of dyad observations.

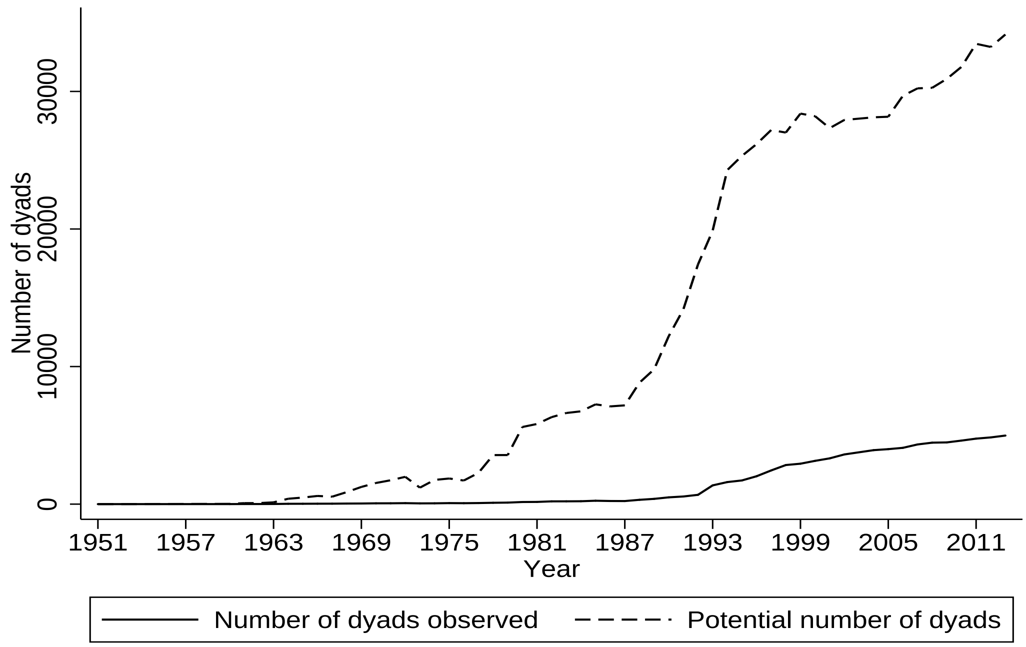

Figure 1.4: Number of dyad observations by year

1.4 The gravity model

The gravity model was originally introduced by Tinbergen (1962) to study international trade flows. Gravity models are used to model other forms of international factor movements: Foreign Direct Investment (FDI) (e.g., Stone and Bang (1999)), trade and FDI (e.g., Bergstrand and Egger (2007)) and migration (e.g., Vanderkamp (1977). Leamer and Levinsohn (1995) claim that gravity models have produced some of the clearest and most robust empirical findings in economics.

Anderson (2011) argues that the gravity model has been *"one of the most successful empirical models in economics, ordering remarkably well the enormous observed variation in economic interaction across space in both trade and factor movements"*. However, gravity models did not have a significant impact on the subject of international economics until they were justified on theoretical grounds (e.g., Anderson and van Wincoop (2003)).

When applied to international trade in goods, the economic analogy to Newton's Law of

Gravitation is described as follows. Omitting reference to time, we have that the flow of goods between a country of origin i and a destination country j, M_{ij}, is proportional to the product of the masses of the origin and destination countries, Y_i and Y_j, and inversely proportional to the squared distance between them, D_{ij}, i.e. $M_{ij} = Y_i Y_j / D_{ij}^2$, where the masses are proxied by the GDP of the two countries.

The gravity model has also been applied to describe forced migration at least once (see Iqbal (2007)). Note that the gravity models of forced and voluntary migration use the counts of refugees and migrants as dependent variables, respectively. Thus, the dependent variable is a *stock,* while the gravity models of trade and foreign investment use *flow*s. We will stick to the norm in the literature and use the stock variable as the outcome of interest. When trying to explain forced migration, the mass variable of the gravity model is population.

The population of the country of origin measures the size of the potential number of people "at risk" while the population of the destination country measures its capacity to host refugees. The square distance is a proxy of the cost that refugees face when moving from the origin to the asylum country.

Two key factors generating refugee populations are the presence of armed conflict, and the level of civil liberties of both the origin and asylum countries. On the one hand, countries with low standards of civil liberties will tend to "produce" more refugees, while countries with high levels of civil liberties ought to be more willing to host forced migrants.

On the other hand, people will escape from countries in conflict and will be reluctant to enter countries in conflict. In addition, other potential determinants of forced migration include dyadic variables such as whether the origin and asylum countries share a common language or a border.

Including civil liberties, conflict and other potential determinants of forced migration, we can write an extended gravity equation where the stock of refugees of origin country i hosted by country j, R_{ij}, is

$$R_{ijt} = \kappa \left(\prod_{k=1}^{K} X_{kit}^{\alpha_k} \prod_{l=1}^{L} X_{ljt}^{\beta_l} \prod_{s=1}^{S} Z_{sijt}^{\gamma_s} \right) \vartheta_i \delta_j \tag{1.1}$$

where X_{ki}, $k = 1,...,K$, are origin-specific variables, X_{lj}, $l = 1,...,L$, are destination-specific variables, Z_{sij}, $s = 1,...,S$, are variables that vary with both origin and destination, and ϑ_i and δ_j are time-invariant origin- and destination-specific unobserved factors. Note that all variables are now weighted by a corresponding parameter and the RHS includes a gravitational constant, κ.

1.5 Estimation methods

Estimation of gravity equations have led authors to two debates. The first debate deals with the specification of the gravity equation in logarithmic or exponential form. Most scholars use a convenient log-linearized transformation of Equation 1.1, which is assumed to hold on average, leading to the conditional mean regression equation:

$$E\left(\ln R_{ijt} \mid X_{it}, X_{jt}, Z_{ijt}, \vartheta_i, \delta_j\right) =$$

$$= \ln\delta + \sum_{k}^{K} \alpha_k \ln X_{kit} + \sum_{l}^{L} \beta_l \ln X_{ljt} + \sum_{s}^{S} \gamma_s \ln Z_{sijt} + \ln\vartheta_i + \ln\delta_j \tag{1.2}$$

which is then estimated by ordinary least squares. This procedure has been criticized by Flowerdew and Aitkin (1982) and Santos Silva and Tenreyro (2006) on several grounds, particularly, the existence of zero values of the dependent variable, which invalidates the logarithmic transformation or forces the researcher to drop observation with zeros in the outcome of interest.

On occasions, researchers take logs after adding a small quantity (typically 1) to the dependent variable, so that the dependent variable takes a strictly positive value and the logarithmic transformation is feasible. Note, however, that such a transformation is incompatible with the gravity model, as one would have to add that small quantity to both the left-hand side (LHS) and right-hand side (RHS) of the gravity equation and then the log-linearization is no longer feasible. Instead, Santos Silva and Tenreyro (2006) suggest modeling the conditional mean of the dependent variable itself as an exponential function of the form

$$E\left(R_{ijt} \mid X_{it}, X_{jt}, Z_{ijt}, \vartheta_i, \delta_j\right) =$$

$$= \exp\left(\ln\delta + \sum_k^K \alpha_k \ln X_{kit} + \sum_l^L \beta_l \ln X_{ljt} + \sum_s^S \gamma_s \ln Z_{sijt} + \ln\vartheta_i + \ln\delta_j\right) \tag{1.3}$$

which can be estimated using Poisson-GMM-like estimators. Note that both the logarithmic and exponential regressions have all regressors measured in logs and therefore the parameters we aim to estimate are the same whether estimated in log-linearized or exponential mean forms. Therefore, the parameters of logarithmic and exponential specifications are directly comparable and can be interpreted as elasticities.

A second debate in the literature involves the issue of whether to account for unobserved heterogeneity in dyadic data modeling. Green et al. (2001) warned against "dirty pooling" and recommended using dyad-specific fixed effects, while Beck and Katz (2001) suggested that using dyad fixed effects is sometimes like "throwing out the baby with the bath water".

Accounting for unobserved heterogeneity is necessary when it is correlated with the covariates; however, conventional fixed effects methods cannot be used to identify the effect of time-invariant covariates, which are often of research interest. In our empirical analysis, we report pooled and fixed effects evidence. In addition, we use the PSM-GMM estimator introduced by Blundell et al. (2002).

This estimation procedure accounts for unobserved heterogeneity by making use of pre-sample information on the dependent variable. The pre-sample average over destinations (origins) is proportional to the unobserved time-invariant origin-(destination-) specific heterogeneity and therefore can be used as aproxy.[6]

Using the observed refugee counts for the 1951-1989 pre-sample period, we calculate the sample averages of those counts for each country of origin and country of asylum/residence and use these averages as proxies for the unobserved origin-specific and destination-specific

[6] In order to see this, average equation 1.1 over destinations (respectively, origins) during the pre-sample period.

unobserved effects. By doing so, we avoid “dirty pooling” and account for unobserved heterogeneity but do not “throw[..] out the baby with the bath water”.

It is worth noting that if gravity model 1.1 holds, there is no need to include interaction terms on the RHS of the logarithmic or exponential regressions. It is also worth noting that, in the empirical part of the chapter, several covariates are categorical, for instance, civil liberties is measured in seven ordered categories. Thus, we approximate the regression function as a linear function of the dummy indicators representing categories.

It was not until very recently that researchers developed methods to adjust standard errors for arbitrary correlation among dyads. Standard errors clustered at the dyad level, often used in dyadic regressions, are robust to correlation among observations within a dyad. The dyadic correlation that the new developments account for is between dyads.

To make the point, consider two observations of a dyadic relation, say pairs (i,j) and (k,l). It could be the case that an observation from dyad (i,j) is correlated with another observation from dyad (k,l) if $i=k$ or $i=l$ or $j=k$ or $j=l$. In words, consider an observation corresponding to the count of refugees from Syria to Germany.

The disturbance corresponding to that observation could be correlated with any other disturbance whose origin or destination is either Syria or Germany. The idea was first developed within network analysis, Fafchamps and Gubert (2007), and extended by Cameron and Miller (2014) and Aronow et al. (2015).

Alternatively, standard errors can be computed using a randomization procedure, as in Erikson et al. (2014), but this might entail a considerably large numerical exercise. In a Monte Carlo experiment, Cameron and Miller (2014) find that dyadic correlation robust standard errors are several times larger than the typically used heteroskedasticity robust standard errors (which are those obtained by allowing for clustering at the dyad level).

As a consequence of the timing of these developments, previous empirical evidence does not employ these recently developed methods, so previously made inferences could be incorrect. In our empirical application, we compute standard errors that account for the so called two-way dyadic correlation, as suggested by Cameron et al. (2011). Under this type of dyadic correlation, disturbances are correlated if they share a common origin or a common destination country.

Based on the previous example, the disturbance corresponding to the Syria-Germany dyad would be allowed to be correlated with any other disturbance from an observation whose origin is Syria or whose destination is Germany. We believe that the two-way dyadic correlation structure is appropriate for refugee count data.

However, should this assumption be incorrect so that the disturbance from the Syria-Germany dyad is also correlated with other disturbances from dyads where Syria is the destination or Germany is the origin, the properly adjusted standard errors would be about ten per cent larger as shown by Cameron and Miller (2014).

A final comment in this methods section deals with the zero-inflated models often used to analyze count data, e.g. Moore and Shellman (2007b). We believe that zero-inflated count models are not appropriate to model refugee counts. These models are often used to analyze data arising from a mixture of two different processes or distributions.

For instance, consider the number of insurance claims from a random sample of individuals. In this case, many of the zeros will correspond to individuals not exposed to the risk, while some of the zeros will correspond to individuals exposed to the risk but who did not claimed any insurance compensation. This type of data is naturally described by a zero-inflated count model. Refugee data, we believe, accrue from only one process.

The number of refugees fleeing a country could be positive or zero, but all countries could potentially have positive numbers and in fact most of them do. Nevertheless, with dyadic data, it turns out that many dyads have zero counts because of great distances between countries or because the number of destination countries is larger than the number of origin countries. The Poisson-type exponential regression model used in this chapter can account for a large number of zeros.

1.6 Empirics

As illustrated in Section 1.3, although refugee counts data run from 1951, it would not be until the 90s when most countries would be both origin and destination; hence, we restrict our analysis to the 1990-2013 period, using 1951-1989 as the pre-sample period referred to above.

We created a dyadic data set as follows. First, we listed all 210 countries that were the origin or destination of at least one refugee in the UNHCR data, excluding some overseas territories and micro-states. Second, we constructed all feasible dyads among the 210 countries. Third, we extended those dyads from 1951 to 2013, excluding the dyads involving newly formed countries for the time periods prior to their inception. Fourth, we merged refugee data and the other covariates into this data set resulting in 76,360 positive refugee counts and 970,055 zero refugee counts. As there are missing covariate values, we ended up running regressions with up to 697,845 observations for 31,507 different dyads from 178 origin and destination countries.

All regressions include the same set of regressors: (Log) Population in the origin and destination countries, dummy variables indicating whether there is an armed conflict in the country of origin or destination, and two sets of dummies that correspond to the (ordered) categories of civil liberties in the origin and asylum/residence countries, from Category 1 (omitted) representing the highest level of civil liberties to Category 7 with the lowest level of civil liberties. In addition, we also include (Log) GDP per capita of the origin and destination countries to allow for the possibility that besides conflict and persecution, refugee counts include economic migrants who flee the origin country and might choose to go to richer destination countries.

Table 1.2 reports estimates of the (log-linearized) gravity equation using a total of (up to) 64,516 directed dyadic positive counts. Each regression is reported in a double column for monadic variables and a single column for dyadic ones. Reported standard errors are adjusted for two-way clustering.

The pooled logarithmic specification suggests that population is a significant determinant of forced migration and that more populated countries tend to be source and destination of larger numbers of refugees. Notice that the parameter estimate of population at the destination country is more than three times the size of the parameter estimate of population at the origin country.

Conflict at the origin country has a positive and significant effect on the number of refugees who flee from that country, but conflict at the destination country is not a significant determinant of which destination they choose. According to this finding, refugees escape from conflict and flee to countries where there might or might not be conflict.

The level of civil liberties, both at the origin and the asylum country, is a significant determinant of refugee flows, albeit with different signs. Civil liberties dummies numbered 2 to 7 have positive coefficients for the country of origin and negative ones for the destination country indicating that, with respect to origin and destination countries with the highest standards of civil liberties, lower standards of civil liberties at the origin (respectively, destination) country tend to increase (decrease) the number of refugees who flee from that source (enter that destination). Looking at the economic motivation, the pooled regression estimates suggest that poorer countries, with lower GDP per capita, tend to be the origin of significantly more refugees, but richer countries do not host significantly more refugees.

All dyadic regressors are statistically significant. Distance between source and destination countries acts as a "transportation cost" reducing the counts of refugees between distant countries. Sharing a border is strongly significant, indicating that destination countries often receive forced migrants from neighbor countries. Sharing a common language also affects the number of refugees flows positively and significantly.

The "Colonizer" variable indicating whether the destination country was a former colonizer of the origin country has a positive and significant effect on refugee counts. The pooled regression estimates occur in "dirty pooling" as they do not account for unobserved heterogeneity.

Ignoring unobserved heterogeneity generates two problems. First, parameter estimates are less efficient than those we would obtain by accounting for such heterogeneity in a random effects regression. Although this is an important issue, it does not seem to be relevant in our analysis as standard errors are already very small.

A second more important problem is that failing to account for unobserved heterogeneity might result in inconsistent parameter estimates if those unobserved factors not only affect the outcome variable (refugee counts), but also affect the covariates included in our regressions. To cope with this problem, Table 1.2 also reports dyad-specific fixed effects. The results are now very different from those obtained in the pooled regression.

Table 1.2: Logarithmic specification

	Pooled			Fixed effects		
	Origin		Destination	Origin		Destination
Log - Population	0.1621***		0.4911***	0.9390*		-0.9646
	(0.0541)		(0.0719)	(0.5260)		(0.6844)
Conflict	0.6271***		0.0773	0.0652		0.1821
	(0.1460)		(0.3561)	(0.0490)		(0.2787)
Civil Liberties_2	1.3163***		-1.5101***	0.8151***		-0.2992**
	(0.2894)		(0.2364)	(0.2679)		(0.1363)
Civil Liberties_3	1.4540***		-1.9193***	1.2496***		-0.2355
	(0.3407)		(0.2999)	(0.3535)		(0.1778)
Civil Liberties_4	1.6242***		-1.8961***	1.3541***		0.0128
	(0.3581)		(0.3000)	(0.3648)		(0.2071)
Civil Liberties_5	1.8572***		-1.8873***	1.5036***		-0.0770
	(0.3576)		(0.3362)	(0.3781)		(0.2447)
Civil Liberties_6	2.2289***		-1.6760***	1.5959***		-0.0991
	(0.3853)		(0.3571)	(0.3794)		(0.2690)
Civil Liberties_7	2.4175***		-1.3430***	1.7728***		0.1627
	(0.4404)		(0.4674)	(0.4277)		(0.3905)
Log - GDP pc	-0.2048**		0.1293	-0.9226***		-0.2351
	(0.0809)		(0.1028)	(0.2513)		(0.4115)
Log - distance		-0.7485***				
		(0.1018)				
Contiguous		2.3492***				
		(0.3674)				
Common language		0.4312**				
		(0.1731)				
Colonizer		0.5833**				
		(0.2732)				
Constant		-6.9359***			11.3882	
		(2.2205)			(15.5573)	
Observations		64,516			64,516	
Countries	178		163	178		163
Dyads		5,868			5,868	
R-squared		0.2818				

All regressions include year dummies. Two-way clustered robust standard errors in parentheses.
One, two and three starts stand for 10, 5 and 1 per cent significance levels.

The population estimated coefficients are much larger than those of the pooled regressions for the origin country. For the destination country, population is a negative but insignificant influence.

Parameter estimates associated with conflict at the origin country are now much smaller and insignificant. Notice that this result goes against the basic intuition that conflict at the origin country is a determinant of refugee flows.

Dummies for civil liberties at the origin country have positive and significant but smaller parameter estimates. Civil liberties at the destination country turn out to be insignificant, with the exception of the dummy for the second level of civil liberties.

Therefore, civil liberties in countries having the second highest civil liberties standards would receive fewer refugees than those in the reference group, while countries with low standards of civil liberties would not host necessarily more refugees.

GDP per capita at the origin country has a positive and significant effect, while it is insignificant at the destination country, as it was in the pooled regression. Finally, the fixed effects estimation method fails to identify the effect of time-invariant dyadic variables.

Why are the pooled and fixed effects results different? Pooling all observations together and ignoring unobserved heterogeneity could be biasing the results in the "dirty pooling" regression if unobserved heterogeneity is correlated with the regressors.

On the other hand, the dyad-specific fixed effects regression cannot identify the effect of the time-invariant covariates.

The dyad-specific fixed effects method does identify the parameters associated with time-varying regressors, but because civil liberties tend to vary very little and civil liberties for some countries remain unchanged during the entire sample period, the identification strategy turns out to have very little power.

In a hypothetical limiting case when civil liberties remain constant over the sample period, the dyad-specific fixed effects method would fail to identify the effect of civil liberties on refugee migration.

Therefore, as the civil liberties dummies do not vary much over time, their effect on refugee migration counts can be difficult to identify in general, and especially when the association between refugee migration counts and civil liberties is likely to be weaker for the destination countries than for origin countries.

Table 1.3 reports the exponential regression estimates. The pooled Poisson-GMM estimates use the 64,516 observations used in the pooled logarithmic regression in Table 1.2 plus 633,329 observations with zero counts for a total of 697,845 observations.

Compared with the pooled logarithmic regression, the pooled exponential regression estimates are similar in terms of statistical significance.

The exponential regression includes many zero count observations and therefore the reference point, with respect to which dummy variable coefficients are to be interpreted, is different from that of the logarithmic regression.

This may explain why the pooled exponential regression results in larger parameter estimates for conflict and civil liberties at the origin country.

Table 1.3 reports fixed effects Poisson estimates. As far as we know, there is no known method to adjust standard errors of the fixed effects Poisson maximum likelihood estimation method for dyadic correlation patterns, thus, we report bootstrapped standard errors.

The fixed effects Poisson method restricts the analysis to all those dyads with at least one positive count during the sample period. Thus, it restricts the sample depending on the values of the dependent variable.

As in the logarithmic regression, the Poisson fixed-effects method fails to identify the effect of time invariant covariates, and for those covariates that vary little, like the conflict indicators, the identification strategy is very weak.

As a result, conflict at the origin country is not a significant determinant of refugee counts, a finding that goes against intuition and previous empirical evidence. Similarly, some of the civil liberties dummies, both at the origin and destination countries, lose significance.

Table 1.3 also reports the PSM-GMM estimates. PSM-GMM estimates account for unobserved heterogeneity by using pre-sample averages of the dependent variable as proxies for the unobserved heterogeneity.

Although it was not originally designed for dyadic data, we tailor it for use with dyadic data.

In particular, we use the pre-sample average of the number of refugees with a specific origin country during the 1951-1989 period, prior to the estimation period, as a proxy for the unobserved origin-specific heterogeneity.

Similarly, we use the pre-sample mean of the number of refugees with a specific destination country during the 1951-1989 period as a proxy for the unobserved destination-specific heterogeneity. These proxy measures of origin and destination-specific sources of heterogeneity are then included as regressors in the exponential mean equation.

As the pre-sample means enter the exponential function in logarithms, the sample is restricted to the 244,710 observations with positive pre-sample means.

Therefore, the PSM-GMM estimation does not include observations on countries with no pre-sample refugee counts (either as origin or destination) including newly created countries that did not exist during the pre-sample period.

Note, however, that the PSM-GMM estimator restricts the sample in a more natural way than the logarithmic regression and the Poisson fixed effects estimator. It restricts the estimation sample to those dyads corresponding to origin and destination countries with at least one positive refugee count during the period prior to the estimation sample, allowing for zero counts during the estimation sample.

Thus, censoring occurs depending on the value of a predetermined variable. Instead, the logarithmic regression constrains the sample to those dyad observations with positive refugee counts during the sample period. That is, censoring occurs depending on the value of the endogenous variable.

Table 1.3: Exponential mean specification

	Pooled-Poisson			Fixed Effects Poisson		PSM-GMM		
	Origin		Destination	Origin	Destination	Origin		Destination
Log - Population	-0.0084		0.5233***	-2.3180*	1.4759	0.0297		0.3710***
	(0.0880)		(0.0864)	(1.3820)	(1.0747)	(0.1150)		(0.0744)
Conflict	1.2041***		0.3905	0.2457	-0.0981	0.8361***		-0.1179
	(0.2938)		(0.3371)	(0.1672)	(0.1119)	(0.2828)		(0.1705)
Civil Liberties_2	3.3799***		-2.0900***	0.9003**	-1.5693***	2.6073***		-1.9438***
	(0.5493)		(0.2898)	(0.3723)	(0.4619)	(0.5572)		(0.4555)
Civil Liberties_3	3.6004***		-2.2667***	0.6259	-1.4911***	3.1302***		-1.5686***
	(0.5608)		(0.4419)	(0.5157)	(0.5589)	(0.4918)		(0.3647)
Civil Liberties_4	4.2982***		-1.9258***	1.2112**	-1.1585**	3.8361***		-1.7197***
	(0.6131)		(0.4267)	(0.5556)	(0.5684)	(0.5325)		(0.3084)
Civil Liberties_5	4.7430***		-1.6836***	1.4459**	-0.8734	3.9548***		-1.7895***
	(0.5730)		(0.4730)	(0.6620)	(0.6683)	(0.5236)		(0.3461)
Civil Liberties_6	5.4198***		-1.6599***	1.6382**	-0.6544	4.3752***		-1.8666***
	(0.6146)		(0.4891)	(0.6662)	(0.7345)	(0.5056)		(0.0832)
Civil Liberties_7	5.9283***		-1.8781***	1.3129	-0.3986	4.7791***		-1.8465***
	(0.7693)		(0.4766)	(0.8179)	(0.8153)	(0.5650)		(0.1430)
Log - GDP pc	-0.5525***		0.2130	-0.9758***	0.4051	-0.1970		0.2371*
	(0.1649)		(0.2505)	(0.2167)	(0.3505)	(0.1422)		(0.1406)
Log - distance		-0.5945***					-0.6016***	
		(0.1243)					(0.1135)	
Contiguous		4.1435***					4.2221***	
		(0.5397)					(0.4180)	
Common language		-0.3577					-0.2150	
		(0.2318)					(0.3756)	
Colonizer		-0.7459					-0.4637	
		(0.7071)					(0.6278)	
Pre-sample Mean						0.1835***		0.3851***
						(0.0572)		(0.0921)
Constant		-4.5608					-6.7238***	
		(3.4847)					(2.5034)	
Observations		697,845		133,137			244,710	
Countries	178		178	178	163	97		112
Dyads		31,507		5,868			10,793	
Pseudo R-squared		0.6360					0.7276	

All regressions include year dummies. Two-way clustered robust standard errors in parentheses.
One, two and three starts stand for 10, 5 and 1 per cent significance levels.

The PSM-GMM estimates associated with conflict and civil liberties at the origin country are somewhat smaller than the corresponding pooled Poisson estimates. Statistical significance of PSM-GMM estimates is similar to the pooled Poisson case with the exception that GDP per capita, which turns out to be significant at the destination, but not at the origin country.

As argued in the introduction, one of the contributions of this chapter is to report two-way clustered standard errors as opposed to the previous literature, which uses dyad-clustered standard errors.

For comparison, Tables 1.4 and 1.5 report the same regressions as Tables 1.2 and 1.3, but using dyad-clustered standard errors. Therefore, parameter estimates are identical in the two sets of Tables, they differ in the standard errors.

We first focus on the logarithmic specification. Two-way clustered standard errors are much larger than the dyad-clustered standard errors, and are sometimes up to about three times larger. As a result, some covariates lose statistical significance when assessing significance with two-way clustered standard errors in Table 1.2.

For instance, GDP per capita at the destination country loses significance in the pooled regression with the logarithmic specification and population and conflict at the destination country lose significance in the fixed effects case.

Turning to the pooled exponential regression, two-way clustered standard errors are up to 30 per cent larger in Table 1.3 than in Table 1.5, but in some cases are even smaller than the dyad-clustered standard errors.

It seems that the issue of using two-way clustered standard errors instead of dyad-clustered standard errors is much more important when using the logarithmic specification.

With the exponential specification, statistical significance is almost unchanged qualitatively, only GDP per capita at the destination country becomes marginally significant with the two-way clustered standard errors.

Table 1.4: Logarithmic specification with dyad-clustered standard errors

	(1) Pooled			(2) Dyad Fixed Effects		
	Origin		Destination	Origin		Destination
Log - Population	0.1621***		0.4911***	0.9390***		-0.9646***
	(0.0207)		(0.0251)	(0.2316)		(0.3050)
Conflict	0.6271***		0.0773	0.0652***		0.1821***
	(0.0574)		(0.0935)	(0.0244)		(0.0638)
Civil Liberties_2	1.3163***		-1.5101***	0.8151***		-0.2992***
	(0.1295)		(0.0772)	(0.1348)		(0.0621)
Civil Liberties_3	1.4540***		-1.9193***	1.2496***		-0.2355**
	(0.1411)		(0.1056)	(0.1646)		(0.0931)
Civil Liberties_4	1.6242***		-1.8961***	1.3541***		0.0128
	(0.1381)		(0.1240)	(0.1719)		(0.1099)
Civil Liberties_5	1.8572***		-1.8873***	1.5036***		-0.0770
	(0.1351)		(0.1321)	(0.1768)		(0.1435)
Civil Liberties_6	2.2289***		-1.6760***	1.5959***		-0.0991
	(0.1439)		(0.1915)	(0.1841)		(0.1690)
Civil Liberties_7	2.4175***		-1.3430***	1.7728***		0.1627
	(0.1640)		(0.3061)	(0.2022)		(0.2473)
Log - GDP pc	-0.2048***		0.1293***	-0.9226***		-0.2351
	(0.0290)		(0.0454)	(0.0986)		(0.1767)
Log - distance		-0.7485***				
		(0.0435)				
Contiguous		2.3492***				
		(0.2224)				
Common language		0.4312***				
		(0.0841)				
Colonizer		0.5833***				
		(0.2056)				
Constant		-6.9359***			11.3882*	
		(0.7644)			(6.4068)	
Observations		64,516			64,516	
Number of countries	178		163	178		163
# dyads		5,868			5,868	
R-squared		0.2818				

All regressions include year dummies. Standard errors clustered at the dyad.
One, two and three starts stand for 10, 5 and 1 per cent significance levels.

Table 1.5: Exponential mean specification with dyad-clustered standard errors

	(1)			(2)		
		Pooled			Pre-Sample Mean GMM	
	Origin		Destination	Origin		Destination
Log - Population	-0.0084		0.5233***	0.0297		0.3710***
	(0.0697)		(0.0851)	(0.0854)		(0.0998)
Conflict	1.2041***		0.3905	0.8361***		-0.1179
	(0.2357)		(0.2577)	(0.2242)		(0.1910)
Civil Liberties_2	3.3799***		-2.0900***	2.6073***		-1.9438***
	(0.4813)		(0.3662)	(0.5591)		(0.2865)
Civil Liberties_3	3.6004***		-2.2667***	3.1302***		-1.5686***
	(0.4882)		(0.4473)	(0.5821)		(0.4396)
Civil Liberties_4	4.2982***		-1.9258***	3.8361***		-1.7197***
	(0.4909)		(0.4308)	(0.5598)		(0.3913)
Civil Liberties_5	4.7430***		-1.6836***	3.9548***		-1.7895***
	(0.4489)		(0.4048)	(0.5565)		(0.3830)
Civil Liberties_6	5.4198***		-1.6599***	4.3752***		-1.8666***
	(0.4521)		(0.5598)	(0.5538)		(0.4825)
Civil Liberties_7	5.9283***		-1.8781***	4.7791***		-1.8465***
	(0.5957)		(0.5045)	(0.5825)		(0.4346)
Log-GDP pc PPP	-0.5525***		0.2130	-0.1970		0.2371
	(0.1303)		(0.1999)	(0.1327)		(0.1487)
Log - distance		-0.5945***			-0.6016***	
		(0.1421)			(0.1144)	
Contiguous		4.1435***			4.2221***	
		(0.3936)			(0.3634)	
Comm. language		-0.3577			-0.2150	
		(0.3968)			(0.3470)	
Former colonizer		-0.7459			-0.4637	
		(0.5107)			(0.4332)	
Pre-Sample Mean				0.1835***		0.3851***
				(0.0424)		(0.0948)
Constant		-4.5608			-6.7238**	
		(2.9263)			(2.8628)	
# observations		697,845			244,710	
# countries	178		178	97		112
# dyads		31,507			10,793	
Pseudo-R^2		0.6360			0.7276	

All regressions include year dummies. Standard errors clustered at the dyad in parentheses.
One, two and three starts stand for 10, 5 and 1 per cent significance levels.

1.7 Robustness checks

A first robustness check involves how we measure conflict and civil liberties. In other words, we wonder whether the results obtained are robust to the way we measure the covariates. The binary conflict indicator bundles together conflict episodes of varying magnitude or intensity.

To see how conflict intensity affects the number of refugees who leave their country of origin or go to a destination country, we replace the binary indicator with a measure of intensity of each conflict from the Center of Systemic Peace, the ACTOTAL score.

This indicator takes values from 0, which indicates no conflict, 1 meaning a small scale conflict, up to 14, the bloodiest degree of conflict. Some of these categories are empty and some include very few conflict episodes.

Following the procedure explained in Appendix A, we recoded this indicator into nine levels of intensity. Similarly, the civil liberties score, is replaced by a 21-level democracy indicator, the POLITY2 score also from the Center of Systemic Peace. The resulting regressions are too big to be displayed in a Table.

Instead, we report the results in a graph. We only report the PSM-GMM estimates to save space.

Figure 1.5 plots parameter estimates and their ninety-five per cent confidence intervals. Statistical significance of parameter estimates can be easily assessed in the graph by checking if the confidence interval includes the x-axis.

Population at the origin country is not statistically significantly different from zero. The reference or left-out group corresponds to a country of origin and a country of destination experiencing no conflict and having the highest standards of democracy.

Therefore, the positive values of the parameter estimates associated with the conflict intensity dummies indicate higher number of refugees than the reference group.

With one exception, as the intensity of conflict in the origin country increases, the associated coefficients also increase, although not uniformly, and are significant. All the autocracy/democracy dummies, corresponding to the origin country, have positive and significant parameter estimates.

As the level of autocracy decreases or democracy increases, the associated parameter estimates decrease although not uniformly. This indicates that higher standards of democracy at the source country reduce the flow of forced migrants.

GDP per capita at the origin country has a negative and significant effect on the number of refugees that flee that country.This result is in contrast with that reported in Table 1.3 using the PSM-GMM method. When accounting for conflict intensity and a more detailed classification of democracy levels, refugee counts decrease with the GDP per capita of the source country. Population at the destination country has a positive and significant effect. This can mean that populous countries receive more refugees. Two of the conflict intensity dummies for the destination country are negative and significant, and the others are insignificant. Thus, conflict at the destination country does not have a clear effect on refugee migration. All the autocracy/democracy dummies at the destination country have a negative parameter estimate and all but three are statistically significant. Although there is not a uniform pattern for these parameter estimates, less democratic/more autocratic countries tend to host fewer refugees. GDP per capita at the destination countries is statistically insignificant. Distance between origin and destination countries significantly reduces forced migration counts. Sharing a common border significantly increases the flows, while sharing language and the former colonizer indicator do not increase the flows. Overall, the results obtained are similar to those reported in Table 1.3 for the PSM-GMM.

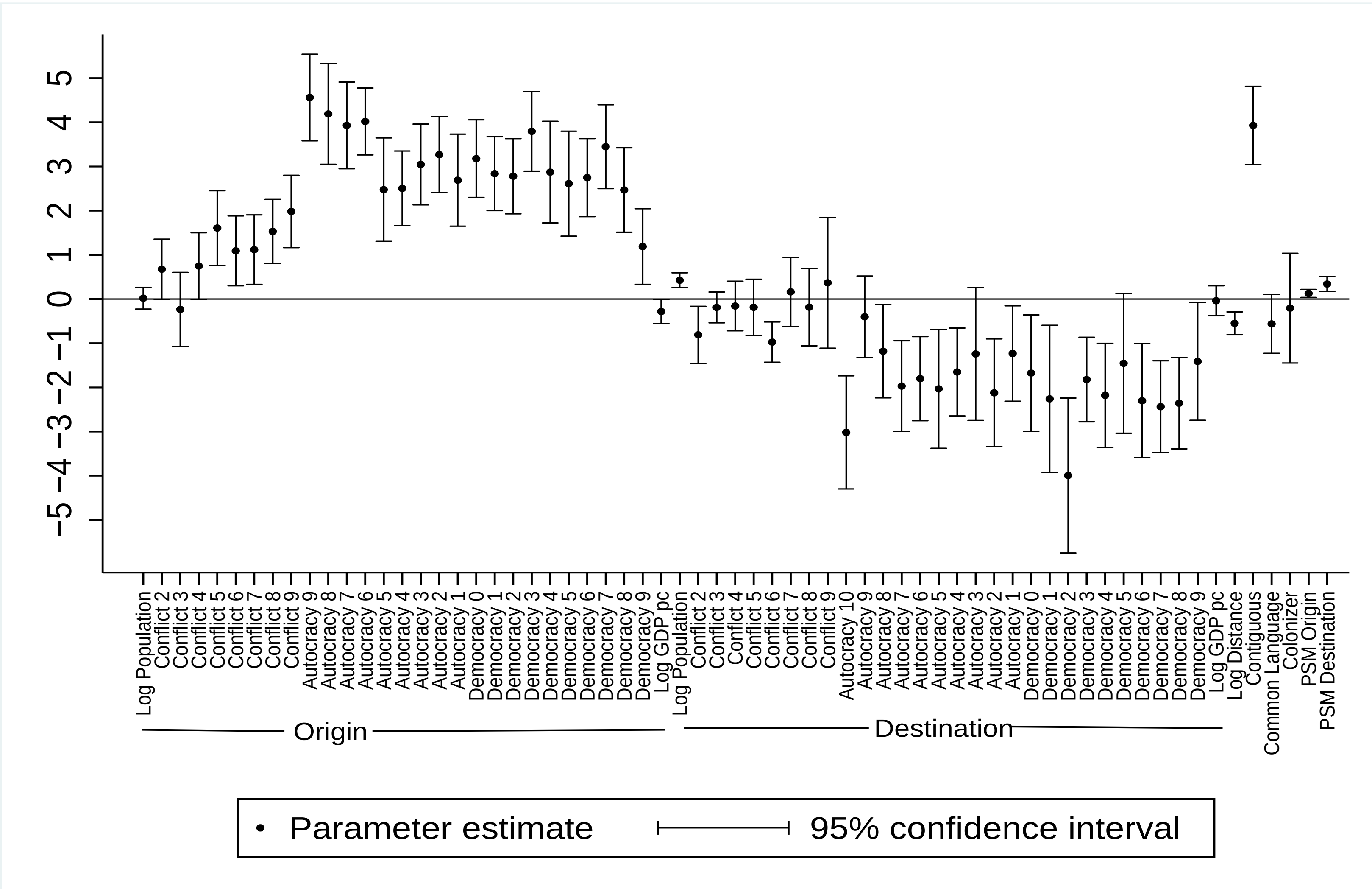

Figure 1.5: Gravity Regression, PSM-GMM estimates with multilevel Conflict and Polity indicators using as references origin and destination countries with no conflict and the highest standards of democracy.

As a second robustness check, we look at how sensible the PSM-GMM estimator is to changing the pre-sample period. Arguably, the pre-sample period used, 1951 to 1989, corresponds to the Cold War, during which the determinants of forced migration may be different from those of the 1990-2013 period used for estimation.

Is the PSM-GMM method robust to using different pre-sample periods? To shed light on this issue, we perform two sensitivity tests, based on two alternative pre-sample periods; from 1951 to 2001, which includes the Cold War but also extends beyond, and from 1992 to 2001, which does not include the Cold War period. Thus, the estimation sample is shortened in both cases to the 2002-2013 period. The results of this sensitivity analysis are reported in Table 1.6. The PSM-GMM estimator restricts the analysis to those countries for which there is at least one positive count during the pre-sample period.

As the two pre-sample periods in Table 1.6 include the 1990s, when the number of dyads with positive counts increased significantly, the results reported in Table 1.6 include more countries, dyads and observations than those reported in Table 1.3 for the PSM-GMM.

Qualitatively, the results reported in Table 1.6 are very similar to those previously reported in Table 1.3 for the PSM-GMM, the only difference is that conflict at the source country is somewhat less significant. We believe that this result is due to the sample period used, rather than the pre-sample mean; the shorter, 2002-2013, sample period includes many more counts of refugees between conflict-free countries, which results in the conflict indicator losing significance. Quantitatively, the results in Table 1.6 are also very similar to those for the sample and pre-sample periods used in the previous section (Table 1.5, PSM-GMM columns).

The only difference is that parameter estimates associated with civil liberties in the origin country are now slightly smaller, while those estimates associated with civil liberties in the destination country are slightly bigger.

Again, we believe this quantitative differences are due to the sample period used, which includes many more positive refugee counts for dyads involving conflict-free countries, thereby assigning a quantitatively more important role to civil liberties.

Table 1.6: Pre-Sample Mean GMM estimates for different pre-sample periods

Pre-Sample Period		1951-2001			1992-2001	
	Origin		Destination	Origin		Destination
Log - Population	0.0960		0.3861***	0.1379		0.3749***
	(0.1098)		(0.0875)	(0.0961)		(0.0806)
Conflict	0.7088**		0.0670	0.5485*		0.0129
	(0.3139)		(0.2535)	(0.2976)		(0.2396)
Civil Liberties_2	2.4864***		-2.0911***	2.4483***		-2.0750***
	(0.6021)		(0.5665)	(0.6056)		(0.5365)
Civil Liberties_3	2.5637***		-1.7160***	2.2384***		-1.8335***
	(0.5773)		(0.4718)	(0.5875)		(0.3996)
Civil Liberties_4	3.2428***		-1.6460***	2.9552***		-1.6709***
	(0.7215)		(0.3308)	(0.6940)		(0.3152)
Civil Liberties_5	3.1054***		-1.7351***	2.6967***		-1.8364***
	(0.5383)		(0.4002)	(0.5409)		(0.4080)
Civil Liberties_6	3.4478***		-2.0422***	3.0200***		-2.2793***
	(0.5016)		(0.2518)	(0.4927)		(0.3234)
Civil Liberties_7	4.0070***		-2.5257***	3.4270***		-2.3537***
	(0.5626)		(0.5359)	(0.5672)		(0.6314)
Log - GDP pc	-0.0084		0.1723	0.0014		0.1265
	(0.1713)		(0.1232)	(0.1464)		(0.1205)
Log - distance		-0.5753***			-0.5357***	
		(0.1545)			(0.1504)	
Contiguous		3.6455***			3.5549***	
		(0.4533)			(0.4209)	
Common language		0.6021			0.6861	
		(0.4604)			(0.4460)	
Colonizer		-0.6612			-0.5870	
		(0.6538)			(0.6255)	
Pre-Sample Mean	0.4632***		0.3396**	0.5645***		0.3748***
	(0.0815)		(0.1403)	(0.0892)		(0.1374)
Constant		-10.0102***			-10.9436***	
		(3.3202)			(3.1059)	
Observations		317,384			313,322	
Countries	171		157	171		155
Dyads		26,692			26,352	
Pseudo R-squared		0.7609			0.7763	

The estimation period is in this Table shortened to 2002-2013. All regressions include year dummies.
Two-way clustered robust standard errors in parentheses.
One, two and three starts stand for 10, 5 and 1 per cent significance levels.

1.8 Discussion and conclusions

This chapter contributes to the literature on forced migration in various respects. First, we use a gravity model to analyze global force migration. Second, we analyze the persons of concern data set from the UNHCR, a database that has previously received very little attention.

Third, we contribute to various methodological debates. On the one hand, we analyze the effect of omitting zero forced migration counts. In our research, including the zeros is tantamount to using an exponential mean specification and omitting them to using a logarithmic specification.

In general, omitting the zeros results in parameter estimates that indicate a smaller impact of some of the explanatory variables on refugee counts. This difference is found to be related to the reference point with respect to which the effect is measured. On the other hand, we contribute to the debate about whether it is more appropriate to account for unobserved heterogeneity via fixed-effects or rely on the pooled specification.

We show how ignoring unobserved heterogeneity might be an example of "dirty pooling" and how accounting for it by means of fixed effects regressions might be like "throwing out the baby with the bath water". The PSM-GMM stands out in this research as a method that can accommodate the presence of zeros and account for unobserved heterogeneity while identifying the effect of time-invariant covariates.

Fourth, the empirical evidence reported in this chapter makes use of the recently developed methods to estimate standard errors that are robust to dyadic correlation. We find evidence suggesting that previous findings based on underestimated standard errors remain after properly adjusting standard errors.

For instance, conflict and absence of civil liberties in the country of origin have been found elsewhere to be fundamental determinants of forced migration and the evidence reported in this chapter using two-way clustered robust standard errors corroborates this finding.

Our evidence also supports the hypothesis that distance and sharing a common border between origin and destination countries are fundamental determinants of forced migration. However, contrary to previously reported evidence, sharing a common language does not seem to influence refugee migration, nor does being a former colonizer influence the count in a significant manner.

Our more substantive finding is the significant influence of the level of civil liberties in the destination country on the number of refugees it receives, a channel not documented before.

Forced migration has lately been in the news as a consequence of the recent exodus from Syria. Thus, there is no need to highlight the importance of this phenomenon and how politicians desperately need to apply policy measures to accommodate such a shock.

The empirical evidence reported in this chapter has some important policy implications. Policy measures to cope with forced migration could be directed to its determinant factors. However, there is little that politicians can do to influence the distance between origin and destination countries or any of the other dyadic determinants, so the only alternative is to influence the incidence of conflict and level of civil liberties or democracy either at the source or destination countries.

For the destination countries, no sane politician would be willing to trade civil liberties or incidence of conflict for a lower inflow of forced migrants. But the risk exists that radicals could direct efforts to reduce the standards of civil liberties to lower the inflow of refugees. Therefore, the only alternative is to fight the refugee problem by making source countries conflict-free with improved standards of civil liberties.

1.9 Appendix A: Data sources of Chapter 1

The data set compiled is a blend of the following data sets:

1. Persons of concern Dataset. United Nations High Commissioner for the Refugees (UNHCR). Time frame: 1960-2014. Variables used: Refugees.
   Availability: http://popstats.unhcr.org/en/time_series

2. Armed Conflict Dataset. Uppsala Conflict Data Program (UCDP) Uppsala University / Peace Research Institute Oslo (PRIO). Time frame: 1946-2014. Variables used: Location. The name(s) of the country/countries whose government(s) have a primary claim to the issue in dispute. Year of observation. The date when the conflict activity reached 25 battle-related deaths in a year. The date when conflict activity ended.
   Availability: https://www.prio.org/Data/Armed-Conflict/UCDP-PRIO/

3. Freedom in the World Dataset. Freedom House. Time frame: 1972-2015. Variables used: Civil Liberties. Availability: https://freedomhouse.org

4. POLITY IV and MAJOR EPISODES OF POLITICAL VIOLENCE (MEPV) Datasets. Center for Systemic Peace. Time frame: 1960-2014. Variables used: POLITY2 (Revised Combined Polity Score). ACTOTAL: Total summed magnitudes of all (societal and interstate). Integer-valued scale ranging from 0 to 14. Transformed into a 0-8 scale by grouping categories (7, 8 and 9) and (10, 11, 12, 13 and 14) of the original scale into categories 7 and 8 of the new one.
Availability: http://www.systemicpeace.org/inscrdata.html

5. Geographical and Distance dataset. CEPII, SciencesPo Department of Economics. Dyadic data set. Variables used: Contiguous countries indicator. Simple distance (most populated cities, km). Common official language indicator. Colonizer: dummy variable which takes value equal to 1 if country of destination has ever been colonizer of the country of origin (a transformation of the original colony variable).
Availability: http://econ.sciences-po.fr/staff/thierry-mayer

6. World Development Indicators. The World Bank. Population. GDP per capita, PPP (current international dollars).
Availability: http://databank.worldbank.org/data/home.aspx

Chapter 2

A spatial model of internal displacement and forced migration

Abstract

This chapter develops a spatial model of internal and external forced migration. We propose a model reminiscent of Hotelling's spatial model in economics and Schelling's model of segregation. Conflict is modeled as a shock that hits a country at certain location and generates displacement of people located near the shock's location. Some displaced people cross a border, thus becoming refugees, while others remain as Internally Displaced Persons (IDPs). The model delivers predictions about how the fractions of a country's population that become refugees and IDPs ought to be related with the intensity of the shock, country size, terrain ruggedness and the degree of geographical proximity of the country with respect to the rest of the world. The predictions of the model are then tested against real data using a panel of 200 countries covering the period 1960-2016. The empirical evidence is broadly in line with the predictions of the model.

2.1 Introduction

Forced displacement occurs when a group of people are obliged to leave their home location unwillingly. Instead, economic migrants choose to leave their home in search of economic opportunities. Forcibly displaced persons can be classified into several groups depending on the root cause of displacement and whether they cross an international border or not. This chapter focuses on two groups of forcibly displaced persons: refugees and Internally Displaced Persons (IDPs).

According to the 1951 United Nations High Commissioner for Refugees (UNHCR) Refugee Convention and the 1967 Protocol (UNHCR 2011), a refugee is someone who "owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality, and is unable to, or owing to such fear, is unwilling to avail himself of the protection of that country."

The 1969 Organization of African Unity Convention (UNHCR 2006) extended the definition to include those who escape from origin countries because of "acts of external aggression, occupation, domination by foreign powers or serious disturbances of public order."

The UNHCR defines IDPs as those "...who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or in order to avoid the effects of armed conflict, situations of generalized violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognized State border" according to the original definition agreed upon at the United Nations in 1998, e.g. Kälin (2008).

According to these definitions, forced displacement can be caused by armed conflict, lack of civil liberties or political rights and natural disasters. However, internal and external displacement share a common cause: armed conflict. When we focus on displaced people who escape from armed conflict, the distinguishing feature between refugees and IDPs is whether they cross an international border.

As refugees and IDPs have a common cause it seems logical to study both groups of people simultaneously. However, the literature on forced displacement is mainly divided into refugee studies on the one hand and internal displacement studies on the other.

In this chapter, we develop a theoretical model where armed conflict is the common cause of internal and external forced migration.

We propose a spatial model of internal displacement and forced migration which has no predecessor in the literature of forced migration. Our model is reminiscent of Hotelling's spatial market model (Hotelling, 1929) and Schelling's model of segregation (Schelling, 1971).

As in Hotelling's model, population is uniformly distributed in a segment of the real line. As in Schelling's model, people change their location according to the dynamics of the model. The spatial model we develop represents armed conflict as a shock that takes place at a particular location and generates a migration flow.

Some displaced people cross a border, thus becoming refugees, while others remain as IDPs. We abstract from other known causes of internal and external forced migration, such as lack of civil liberties or political rights, economic development and natural disasters.[1]

The model predicts how the number of refugees and IDPs, as a fraction of a country's population, varies with the intensity of armed conflict and geographical covariates such as country size, orography and distance to other countries.

In our model, armed conflict has a causal effect on displacement, and the geographical variables exert an effect modification on displacement.

One of the effect modifiers we study is physical country size, i.e. area. Basic intuition indicates that the area of a country must necessarily be a determinant in the fraction of the population affected by a conflict of a particular magnitude.

For instance, in a very large country, a conflict might generate a certain amount of internal displacement of affected people to other parts of the country and no refugees.

However, a conflict of similar magnitude in a sufficiently smaller country would generate

[1]Although the model focuses on conflict-induced displacement, it could describe displacement induced by natural disasters. The only difference would be that those who cross a border would be externally displaced, but according to the definitions given above, would not be refugees.

an outflow or refugees to neighboring countries. A second geographical covariate of interest is the orography. Intuition suggests that displacement should be easier the flatter and obstacle-free a country is, while displacement ought to be much more difficult in a highly mountainous country.

Thus, as a result of a conflict of a particular magnitude, the rougher the orography, the less the displaced people ought to move, hence potentially generating a lower number of refugees. A third geographical variable analyzed is distance to neighboring countries.

Intuition suggests that the more distant neighboring countries are, the lower the number of refugees while total displacement should be independent of the distance. Summarizing, in the model, armed conflict generates displacement while country size, orography and distance to other countries determine the division of displacement into refugees and IDPs.

We develop a barebones model intentionally, with the aim of keeping the mathematical analysis as simple as possible. Due to its simplicity, the model allows for a number of extensions at the cost of more complicated analysis.

In addition, this chapter contributes empirical evidence by testing the predictions of the model against real data. The empirics use a panel data set of 200 countries covering the period 1960-2016. This data set includes country of origin refugee and IDP counts, conflict and geographical data.

We exploit the cross country variation in the geographical variables to analyze their effect modification. In addition, we also account for non-spatial variables which our model ignores, but the literature has found relevant, such as the lack of civil liberties or political rights and Gross Domestic Product (GDP) per capita. The evidence obtained is in line with the model's predictions.

The next section reviews the literature and indicates the contribution of this chapter to it. Section 2.3 develops a spatial model of internal and external displacement. Section 2.4 describes the empirical methods, the data used and reports the empirical evidence. Section 2.5 sums up the main conclusions.

2.2 Literature review and our contribution

The literature on forced displacement is divided into two areas: refugee and IDP studies. A fundamental reason for this divide in the literature is data availability. Refugee destination countries keep track of asylum applications which generate data which can be used to analyze the choice of destination, e.g. Havinga and Böcker (1999), Neumayer (2005a) and Neumayer (2005b).

Other studies analyze dyadic forced displacement flows between countries, e.g. Iqbal (2007), Moore and Shellman (2007b), Barthel and Neumayer (2015) and Echevarria and Gardeazabal (2016).

A stream of the literature builds on the consequences of refugee inflows in the host country. Some explore the role of refugee flows in spreading conflict in the host country e.g. Salehyan and Gleditsch (2006), Salehyan (2008), Milton et al. (2013), Rüegger (2019), Fisk (2019) and Böhmelt et al. (2019).

Another arm focuses on the effect of refugee inflows on the host country's contribution to Official Development Assistance (ODA) (e.g. Czaika, 2009b), the UNHCR (e.g. Roper and Barria, 2010) or UN peacekeeping missions (e.g. Uzonyi, 2015).

All previous references use cross-country data (either longitudinal or dyadic) to analyze external displacement. However, a large number of articles are country case studies. Following Card's (1990) seminal article on the labor market consequences of a large immigration shock, a literature has emerged, e.g. Borjas (2017), Peri and Yasenov (2019, forthcoming).

Similarly, large refugee inflows have prompted research analyzing their effect on the host country's food prices, wealth, labor market, health, education and other outcomes, e.g. Alix-Garcia and Saah (2009), Baez (2011), Tumen (2016), Ceritoglu et al. (2017), Esen and Oğuş Binatlı (2017), Akgündüz and Torun (2018) and Verme and Schuettler (2019).

Relatively less work has been done on the analysis of refugee outflows from origin countries. Davenport et al. (2003) analyzes the determinants of aggregate stocks of migrants, whether internal or external, from origin countries.

More recently, Dreher et al. (2019) analyze the effect of aid receipts on total refugee outflows and also on flows to donor countries. Our chapter analyzes refugee outflows, thus contributing to this less prolific arm of the literature.

When we turn to the literature on internal displacement, we find country case studies. As in refugee studies, there are quantitative assessments of the effect of displacement on host communities. The massive conflict-induced internal displacement in Colombia has been the focus of several articles.

Engel and Ibáñez (2007) and Ibáñez and Vélez (2008) study the determinants of displacement. Morales (2018) studies the impact of displacement on wages in the host communities. Depetris-Chauvin and Santos (2018) analyze the effect of internal displacement on rental prices in host cities.

McEniry et al. (2019) investigate the effect of exposure to displacement on older adults health. Displacement determinants in Indonesia are analyzed in Czaika and Kis-Katos (2009). Kondylis (2008) conducts a resettlement policy evaluation in Rwanda.

Kondylis (2010) assesses the labor market effect of displacement in Bosnia and Herzegovina.

Alix-Garcia et al. (2012) study the price responses to internal displacement and aid in Sudan. Alix-Garcia et al. (2013) estimate the effect of conflict-induced internal displacement on spatial changes in land use.

The spread of conflict has also been studied in relation with internal displacement, e.g. Bohnet et al. (2018).

Other country case studies analyze not only conflict-originated displacement but also internal displacement caused by natural disasters and economic development e.g. Lanjouw et al. (2000), Muggah (2003) and Fernandes (2017).

There are also a few policy analyses Goswami (2007), Lischer (2008), Crisp (2010) and Munive (2019), and also applications of the interview-research methodology, e.g. Ayata and Yukseker (2005).

However, to the best of our knowledge, there is not a single cross-country study on internal displacement. We carry out a cross country empirical analysis of forced internal displacement, thus contributing evidence on this previously unexplored angle.

Theoretical studies of displacement are scant in the literature. A notable exception is Czaika (2009b) who models refugee migration decisions and studies the distribution of burdens from forced migration across countries. Therefore, our model is a contribution in a field with scarce production of theoretical models.

Furthermore, the model designed below builds a bridge between internal and external forced migration. As far as we know, simultaneous modeling of internal and external forced migration has not been attempted before.

Indeed, armed conflict generates displacement. Ball et al. (2002) is perhaps the single most clear example showing the relation between conflict and displacement. Their study of the 1998-1999 conflict in Kosovo shows how displacement counts mirror conflict intensity measured as the number of people killed.

Our model takes the relation between conflict and displacement as granted and analyzes how geographical variables affect displacement, both internally and externally. The literature includes previous studies where geographical factors play a role in determining displacement flows.

In particular, the influence of distance between countries has been incorporated into empirical models of dyadic refugee counts, either through spatial dependence as in Barthel and Neumayer (2015) or directly as in the gravity models of Iqbal (2007) or Echevarria and Gardeazabal (2016). However, the roles of country size and ruggedness of terrain have not been explored before.

2.3 A spatial model of forced migration

For expositional purposes, first we analyze a baseline version of the model with no spatial variables involved in subsection 2.3. This version of the model is the easiest possible and allows us to understand its mechanics. Then, subsection 2.3.1 develops a more sophisticated version of the model where we include the spatial variables and see how they interact with conflict in the determination of displacement.

The baseline model

This section develops a spatial model to explain conflict induced displacement. Three countries, A, B and C, align in the real line. Country A is located to the left of the the origin, country B is located in the interval $[0,1]$ and country C to the right of unity. Points 0 and 1 are the borders.

We will focus on country B, whose population is uniformly distributed on the unit interval. A shock of size s hits country B at location l_s, a point in the unit interval. Figure 2.1 shows the assumptions made so far.

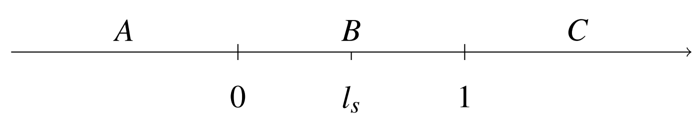

Figure 2.1: Countries A, B and C are located in the real line. Points 0 and 1 are the borders. The shock takes place at location l_s.

The shock affects citizens of country B forcing them to leave their home location if the benefits from staying are lower or equal than the costs. Remaining at their home location yields a benefit b for every citizen, while the cost of staying is

$$c(s, l_i, l_s) = s - |l_i - l_s|, \tag{2.1}$$

where s is the size of the shock, l_i, the location of individual i, and l_s is the location of the shock. In words, the cost of staying equals the size of the shock minus the distance from the location of the individual to the location of the shock.

For simplicity, we normalize the benefit of staying to zero. Under these assumptions, individuals whose distance to the shock location is less than the size of the shock are displaced.

We assume the intensity of the shock, s, ranges from 0 to 1. Notice that a shock whose intensity is equal to 1 would affect the entire population. Therefore, it seems unnecessary to analyze larger shocks. We also assume that there is no international spillover of armed conflict and therefore the shock only affects citizens of country B.

Hence, in those cases where $l_s - s$ is negative or $l_s + s$ is greater than unity, those who decide to move are in the intersection of $[l_s - s, l_s + s]$ and the unit interval. Depending on whether $l_s - s$ is positive or negative and $l_s + s$ is smaller or greater than unity, the proportion of displaced people varies. The four feasible cases are represented by the four regions depicted in Figure 2.2.

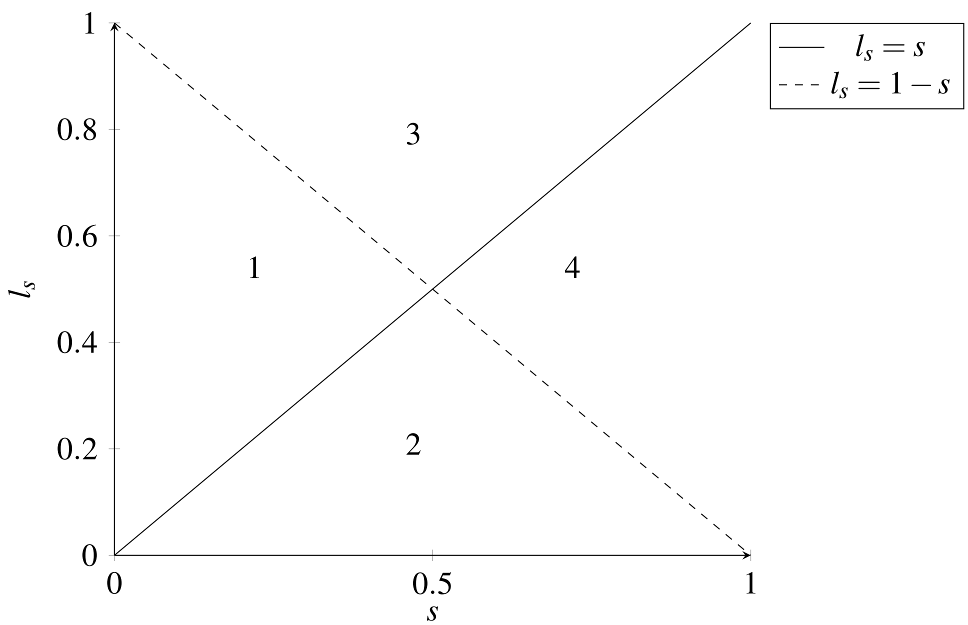

Figure 2.2: Four regions in the shock intensity and shock location space. Total displacement is different in each of these regions. Notice that the two lines cross at $s = 0.5$.

Region 1 corresponds to the case where $l_s - s \geq 0$ and $l_s + s \leq 1$ and represented in Figure 2.3. Under these conditions, the proportion of people displaced equals $2s$, the length of the segment $[l_s - s, l_s + s]$. Within region 2, and as shown in Figure 2.4, $l_s + s \leq 1$ and $l_s - s < 0$, so the shock affects everyone located to the left of l_s and displaced people are a fraction $l_s + s$ of the population. In Region 3, $l_s - s \geq 0$ and $l_s + s > 1$, so displaced people are a fraction $1 - (l_s - s)$ of the population, as captured by Figure 2.5. Finally, in Region 4, $l_s - s < 0$ and $l_s + s >$, so every person is affected and the proportion of displaced people equals 1. Therefore, total displacement, D, is defined over the four regions as follows

$$D = \begin{cases} 2s & if \quad l_s \geq s \quad \& \quad l_s \leq 1 - s \\ l_s + s & if \quad l_s < s \quad \& \quad l_s \leq 1 - s \\ 1 - (l_s - s) & if \quad l_s \geq s \quad \& \quad l_s > 1 - s \\ 1 & if \quad l_s < s \quad \& \quad l_s > 1 - s \end{cases} \tag{2.2}$$

Notice that total displacement is a continuous function of the shock's intensity and location, and non-differentiable at the boundaries between regions. Under the assumptions laid out so far, the model predicts the proportion of the population displaced as a function of shock intensity and location.

With no further assumptions, the model gives no indication as to how far displaced persons move. For simplicity, we assume that displaced persons move a distance equal to the size of the shock, either leftward or rightward depending on which side of the shock location they are.

This form of displacement is consistent with a story where at each point in the line to the left of $l_s - s$ or the right of $l_s + s$, a person not affected by the shock hosts one, and only one, displaced person. As a consequence of people's displacement, some might cross a border becoming refugees while others remain in the home country as IDPs.

More precisely, those people whose distance to a border is less than the size of the shock become refugees. For the time being, we will assume that crossing a border is cost-less and relax this assumption in Section 2.3.1.

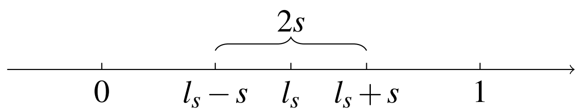

Figure 2.3: A shock of intensity s hits Country B at location l_s. Those in the interval $[l_s - s, l_s + s]$ are affected by the shock. Total displacement is $2s$.

Figure 2.4: A shock of intensity s hits Country B at location l_s. Those in the interval $[0, l_s + s]$ are affected by the shock. Total displacement is $l_s + s$.

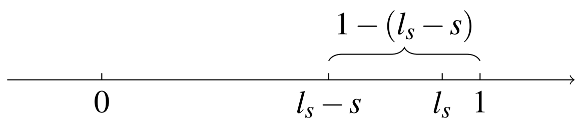

Figure 2.5: A shock of intensity s hits Country B at location l_s. Those in the interval $[l_s - s, 1]$ are affected by the shock. Total displacement is $1 - (l_s - s)$.

Figure 2.6 sorts (s, l_s)-pairs into nine regions depending of their position with respect to four lines.

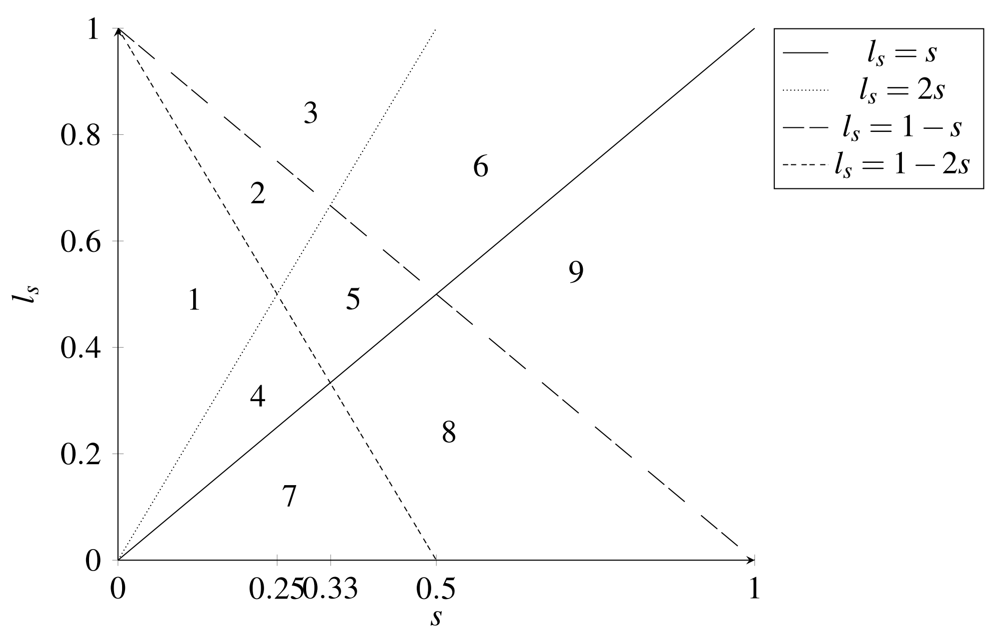

Figure 2.6: Areas of internal and external displacement

Region 1 is further described in Figure 2.7, where the shock affects everyone in interval $[l_s - s, l_s + s]$. Those in the interval $[l_s - s, l_s)$ move to $[l_s - 2s, l_s - s)$ and those initially in the interval $[l_s, l_s + s]$ move to $(l_s + s, l_s + 2s]$.

Because $l_s > 2s$, no one leaves the country through the border with country A, and as $l_s < 1 - 2s$, no one crosses the border with country C. As a consequence, everyone affected by the shock remains internally displaced.

Figure 2.8 represents Region 2 where, as in the previous case, no one reaches the border to Country A, while a fraction $l_s + 2s - 1$ of the population reaches Country C and become refugees.

Figure 2.9 represents Region 3 where all people affected to the right side of l_s, the fraction $1 - l_s$ of the population, become refugees while those affected to the left of l_s remain as IDPs.

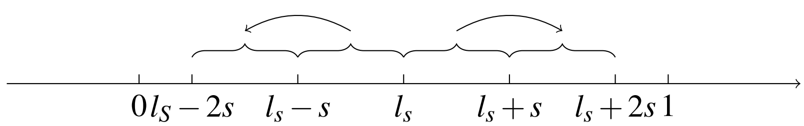

Figure 2.7: A shock of intensity s hits Country B at location l_s. Those in the interval $[l_s - s, l_s)$ move to $[l_s - 2s, l_s - s)$ and those in the interval $[l_s, l_s + s]$ move to $[l_s + s, l_s + 2s]$. There are no flows of refugees between Country B and either Country A or C. All displaced people remain as IDPs.

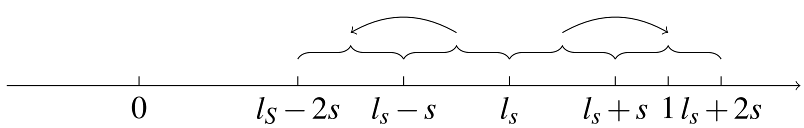

Figure 2.8: A shock of intensity s hits Country B at location l_s. Those in the interval $[l_s - s, l_s)$ move to $[l_s - 2s, l_s - s)$ remaining as IDPs. Those in the interval $[l_s, l_s + s]$ move to $[l_s + s, l_s + 2s]$. A fraction $l_s + 2s - 1$ of the population crosses the border to Country C, thus becoming refugees, and a fraction $1 - l_s - s$ of the population remains as IDPs.

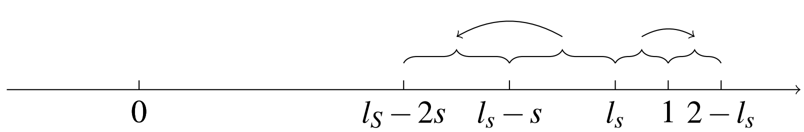

Figure 2.9: A shock of intensity s hits Country B at location l_s. Those in the interval $[l_s - s, l_s)$ move to $[l_s - 2s, l_s - s)$ remaining as IDPs. Those in the interval $[l_s, 1]$ move to $[1, 2 - l_s]$. A fraction $1 - l_s$ of the population crosses the border to Country C, thus becoming refugees.

The reader is invited to carry out the same sort of reasoning for the remaining regions to verify that those who become refugees represent a fraction R of the population given by the following expression,

$$R = \begin{cases} 0 & if \quad 0 < l_s - 2s < l_s - s \quad \& \quad l_s + s < l_s + 2s < 1 \\ l_s + 2s - 1 & if \quad 0 < l_s - 2s < l_s - s \quad \& \quad l_s + s < 1 \leq l_s + 2s \\ 1 - l_s & if \quad 0 < l_s - 2s < l_s - s \quad \& \quad 1 \leq l_s + s < l_s + 2s \\ 2s - l_s & if \quad l_s - 2s \leq 0 < l_s - s \quad \& \quad l_s + s < l_s + 2s < 1 \\ 4s - 1 & if \quad l_s - 2s \leq 0 < l_s - s \quad \& \quad l_s + s < 1 \leq l_s + 2s \\ 2s + 1 - 2l_s & if \quad l_s - 2s \leq 0 < l_s - s \quad \& \quad 1 \leq l_s + s < l_s + 2s \\ l_s & if \quad l_s - 2s < l_s - s \leq 0 \quad \& \quad l_s + s < l_s + 2s < 1 \\ 2l_s + 2s - 1 & if \quad l_s - 2s < l_s - s \leq 0 \quad \& \quad l_s + s < 1 \leq l_s + 2s \\ 1 & if \quad l_s - 2s < l_s - s \leq 0 \quad \& \quad 1 \leq l_s + s < l_s + 2s \end{cases} \tag{2.3}$$

Once we know the proportion of the population displaced and the fraction of the population that refugees represent, the fraction that remain internally displaced is computed as the difference.

This model delivers the fractions of internally and externally displaced populations as functions of the location and the size of the shock.

However, in order to be able to use our model for empirical analysis, it is convenient to integrate out the location variable. With this purpose, we assume the shock location is equally likely to take place at any point in Country B.

Appendix 2.7 shows that, integrating over the shock location, we obtain the conditional expectation of the fraction of people displaced, the fraction externally displaced and, by subtraction, the fraction internally displaced

$$E_{l_s}[D|s] = s(2-s) \quad if \quad s \in [0,1] \tag{2.4}$$

$$E_{l_s}[R|s] = \begin{cases} 2s^2 & if \quad s \in [0, 0.5] \\ \\ 4s - 2s^2 - 1 & if \quad s \in (0.5, 1] \end{cases} \tag{2.5}$$

$$E_{l_s}[I|s] = \begin{cases} s(2-3s) & if \quad s \in [0, 0.5] \\ \\ s^2 - 2s + 1 & if \quad s \in (0.5, 1] \end{cases} \tag{2.6}$$

Notice that, even though equations 2.2 and 2.3 are continuous but not differentiable functions, integrating over the location removes the non differentiability. Figure 2.10 plots the expected values of the proportions of displaced people, refugees and IDPs as continuous and differentiable functions of the shock intensity. In its simplest form, the model delivers the following results. First, both the expected values of the proportions of displaced persons and refugees are increasing functions of the size of the shock. Second, the proportion of IDPs is increasing in the size of the shock for small values of it, has a maximum at $s = 1/3$ and from that point on it is decreasing in the size of the shock.

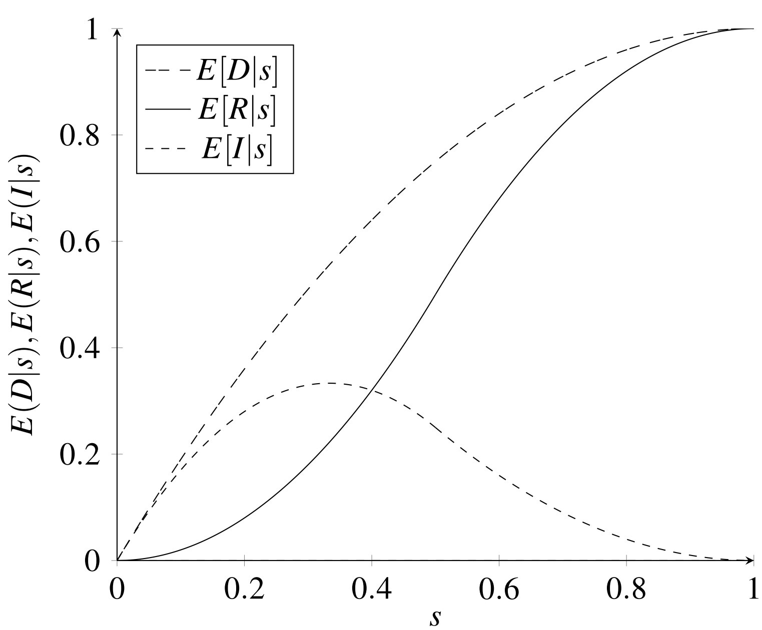

Figure 2.10: Expected values of groups of people

Finally, notice that these functions return fractions of population, thus the total number of people displaced, the total number of refugees and the total number of IDPs can be obtained multiplying these fractions by population. Therefore, population is in this model a natural measure of exposure.

2.3.1 Country size, ruggedness and proximity

The model developed in the previous section assumes the armed conflict shock is the only cause of displacement. In this section, we introduce various spatial variables into our analysis: country size, ruggedness and proximity.

These variables are not causal forces of displacement but they modify the relationship between the shock and displacement and have some bearing in determining how total displacement splits into internal and external.

The area of a country plays an important role concerning internal and external displacement. Intuitively, holding everything else constant, displaced people should find it easier to escape from smaller countries. So far, the model has been developed under the assumption that the size of the Country B is unity.

For simplicity, we assume that the largest country on earth has size unity and the other countries are a fraction a of the largest. Therefore, we now assume that location l_s ranges from zero to a, a value lower than one. Stretching the language, we will refer to a as the "area" of the country, despite the fact our model represents a country as a segment in the real line.

Next we make two assumptions. First, we assume population is uniformly distributed in the $[0,a]$ segment, so the probability density function of population is $1/a$ at each location in $[0,a]$ and zero otherwise. Notice the difference between the standard concept of "population density", say the number of people per square kilometer, and the probability density of the variable population, which describes how population is distributed within the country.

We assume the latter is uniformly distributed but say nothing about the former, that is, the country could have a high or a low "population density". Second, without loss of generality, the size of the shock is now assumed to lay in the interval $[0, a]$. Notice that a shock of size a affects the entire population of the country, so it seems unnecessary to analyze larger shocks.

A geographical factor that affects internal and external displacement is the terrain ruggedness of the source country. Rugged source countries impose an extra difficulty for displacement. Let r denote the degree of ruggedness of a country, a value in the unit interval, with $r = 0$ for a perfectly flat country and $r = 1$ for an abrupt country highly inaccessible.

We assume that in a rugged country, a shock of size s forces people affected to move a fraction γ of the distance displaced in the baseline model, that is, γs, where $\gamma = \gamma(r)$, with $\gamma'(r) < 0$, so the fraction γ is decreasing in ruggedness.

In addition, we assume $\gamma(0) = 1$, so in a perfectly flat country, displaced persons move as described in the baseline model. We also assume that $0 < \gamma(1) < 1$, that is, in a country with the highest ruggedness parameter, $r = 1$, the fraction γ is small but positive.

Distance between countries is another geographical factor of relevance for the analysis of internal and external displacement. Intuitively, holding everything else constant, greater distance between two countries ought to be associated with lower refugee flows.

However, the model analyzed in the previous section does not account for distance between countries. In fact, the model assumes countries A, B and C are contiguous. Let us introduce distance into the analysis. Let d_{ij} be the normalized distance between country i and country j, with $i, j = A, B, C$.

Distance is normalized so that the longest distance between two countries equals unity. However, we continue using the contiguous countries design and model distance between countries as an iceberg-type cost. More specifically, out of all displaced people who reach the border between countries i and j, the fraction that crosses the border is a decreasing function of the distance between countries i and j, $\alpha_{ij} = \alpha\left(d_{ij}\right)$, with $\alpha'\left(d_{ij}\right) < 0$. This fraction equals one when distance is zero, $\alpha(0) = 1$, and is below one and positive when distance between countries i and j is maximal, $0 < \alpha(1) < 1$.

Modeling distance this way abstracts from whether countries are linked by land or necessarily by sea. In addition, it could be argued that this iceberg-type cost associated with crossing a border should be taken into account at the time of deciding whether to move or not.

However, the model assumes that the decision whether to move or not is independent of the cost of crossing a border. The model also abstracts from asylum application costs and the destination country stance towards refugees.

Importantly, notice that the fraction $1-\alpha\left(d_{ij}\right)$ of those who reach the border between countries i and j pile up at the border, in line with a real world observation. Under these assumptions, total displacement is

$$D=\begin{cases} 2\gamma s/a & if \quad l_s-\gamma s\geq 0 \quad \& \quad l_s+\gamma s\leq a \\ \left(l_s+\gamma s\right)/a & if \quad l_s-\gamma s<0 \quad \& \quad l_s+\gamma s\leq a \\ 1-\left(l_s-\gamma s\right)/a & if \quad l_s-\gamma s\geq 0 \quad \& \quad l_s+\gamma s>a \\ 1 & if \quad l_s-\gamma s<0 \quad \& \quad l_s+\gamma s>a \end{cases} \tag{2.7}$$

Similarly, the fraction of country B's population that become refugees is

$$R=\begin{cases} 0 & if \quad 0<l_s-2\gamma s<l_s-\gamma s \quad \& \quad l_s+\gamma s<l_s+2\gamma s<a \\ \alpha_{BC}(l_s+2\gamma s-a)/a & if \quad 0<l_s-2\gamma s<l_s-\gamma s \quad \& \quad l_s+\gamma s<a\leq l_s+2\gamma s \\ \alpha_{BC}(a-l_s)/a & if \quad 0<l_s-2\gamma s<l_s-\gamma s \quad \& \quad a\leq l_s+\gamma s<l_s+2\gamma s \\ \alpha_{AB}(2\gamma s-l_s)/a & if \quad l_s-2\gamma s<0\leq l_s-\gamma s \quad \& \quad l_s+\gamma s<l_s+2\gamma s<a \\ [(2\gamma s-l_s)\alpha_{AB}+(l_s+2\gamma s-a)\alpha_{BC}]/a & if \quad l_s-2\gamma s<0\leq l_s-\gamma s \quad \& \quad l_s+\gamma s<a\leq l_s+2\gamma s \\ [(2\gamma s-l_s)\alpha_{AB}+(a-l_s)\alpha_{BC}]/a & if \quad l_s-2\gamma s<0\leq l_s-\gamma s \quad \& \quad a\leq l_s+\gamma s<l_s+2\gamma s \\ \alpha_{AB}l_s/a & if \quad l_s-2\gamma s<l_s-\gamma s\leq 0 \quad \& \quad l_s+\gamma s<l_s+2\gamma s<a \\ [l_s\alpha_{AB}+(l_s+2\gamma s-a)\alpha_{BC}]/a & if \quad l_s-2\gamma s<l_s-\gamma s\leq 0 \quad \& \quad l_s+\gamma s<a\leq l_s+2\gamma s \\ [l_s\alpha_{AB}+(a-l_s)\alpha_{BC}]/a & if \quad l_s-2\gamma s<l_s-\gamma s\leq 0 \quad \& \quad a\leq l_s+\gamma s<l_s+2\gamma s \end{cases} \tag{2.8}$$

Notice that total displacement depends on how many people are affected by the shock and it does not depend on whether they cross a border or not. Therefore, total displacement does not depend on distance. However, the fraction of country B's population that become refugees does depend on distance.

Appendix 2.7 shows how integrating out the location variable the expected values of total displacement, refugees and IDPs are

$$E_{l_s}[D|s,a,r,d_{AB},d_{BC}] = (2a\gamma s - \gamma^2 s^2)/a^2 \quad if \quad s \in [0,a]$$

$$E_{l_s}[R|s,a,r,d_{AB},d_{BC}] = \begin{cases} \gamma^2 s^2(\alpha_{AB}+\alpha_{BC})/a^2 & if \quad s \in [0,a/2] \\ (2a\gamma s - \gamma^2 s^2 - a^2/2)(\alpha_{AB}+\alpha_{BC})/a^2 & if \quad s \in (a/2,a] \end{cases}$$

$$E_{l_s}[I|s,a,r,d_{AB},d_{BC}] = \begin{cases} [2a\gamma s - \gamma^2 s^2 - \gamma^2 s^2(\alpha_{AB}+\alpha_{BC})]/a^2 & if \quad s \in [0,a/2] \\ [2a\gamma s - \gamma^2 s^2 - (2a\gamma s - \gamma^2 s^2 - a^2/2)(\alpha_{AB}+\alpha_{BC})]/a^2 & if \quad s \in (a/2,a] \end{cases}$$

where the expected values of the relevant fractions are conditional on the size of the shock, area, ruggedness and distances to neighboring countries. Notice that the term $P_B = \alpha_{AB} + \alpha_{BC}$ can be interpreted as a measure of proximity.

Appendix 2.7 shows the derivatives of the expected value of the proportions of refugees, internally displaced and total displacement with respect to area, ruggedness and proximity. The signs of these derivatives indicate that the fraction of the population displaced responds negatively to the area and ruggedness while it is unrelated to proximity.

The proportion of refugees also responds negatively to area and ruggedness, but it is positively related with proximity. However, the signs of the derivatives of the proportion of IDPs with respect to area, ruggedness and proximity depend on the magnitude of the size of the shock, area, proximity and ruggedness in a complex manner.

2.4 The data, empirical methods and evidence

2.4.1 The econometric methods

The dependent variables analyzed are total displacement, total number of refugees and IDPs. We use lagged population as exposure, therefore the dependent variable is the fraction of the population that refugees (respectively, IDPs or total displacement) represent.

The expected value of the fraction of a country's population that become refugees (respectively, IDPs or displaced) is assumed to be an exponential function of a vector of covariates. The exponential regression functions are estimated by the Generalized Method of Moments (GMM), which is numerically identical to the pooled Poisson maximum likelihood estimator. We will term this estimator as Pooled-GMM.

However, the Pooled-GMM estimator does not account for unobserved heterogeneity. Accounting for unobserved time-invariant heterogeneity via country of origin fixed effects cannot simultaneously identify the effect of magnitudes such as area or terrain roughness which are time invariant.

Thus, even though our empirical evidence is based on a panel data set, we cannot resort to using the popular fixed effects methodology, which would preclude inference about time invariant covariates, such as area, ruggedness and proximity. Instead, we use the Pre-Sample Mean Generalized Method of Moments (PSM-GMM) suggested by Blundell et al. (2002).

The PSM-GMM estimation method is like the Pooled-GMM only that it includes an additional regressor: the average value of the dependent variable during a period prior to the sample period.

The pre-sample mean of the dependent variable is used as a proxy for the unobserved heterogeneity. As the dependent variable is measured as a fraction of the population, the pre-sample mean is also measured as a fraction of the population.

2.4.2 The data

Variable definitions and data sources are shown in Appendix 2.6. In this chapter we use refugee and IDP populations from the UNHCR populations of concern database.

Figure 2.11 shows world aggregate time series of the number of IDPs and refugees. Over the last few decades, the number of IDPs has increased remarkably, while the number of refugees has stabilized.

The total number of forcibly displaced people keeps on growing and represents one of the most important problems for humanity. Forced displacement is not a problem of just poor and conflicted countries, it affects almost every country on earth.

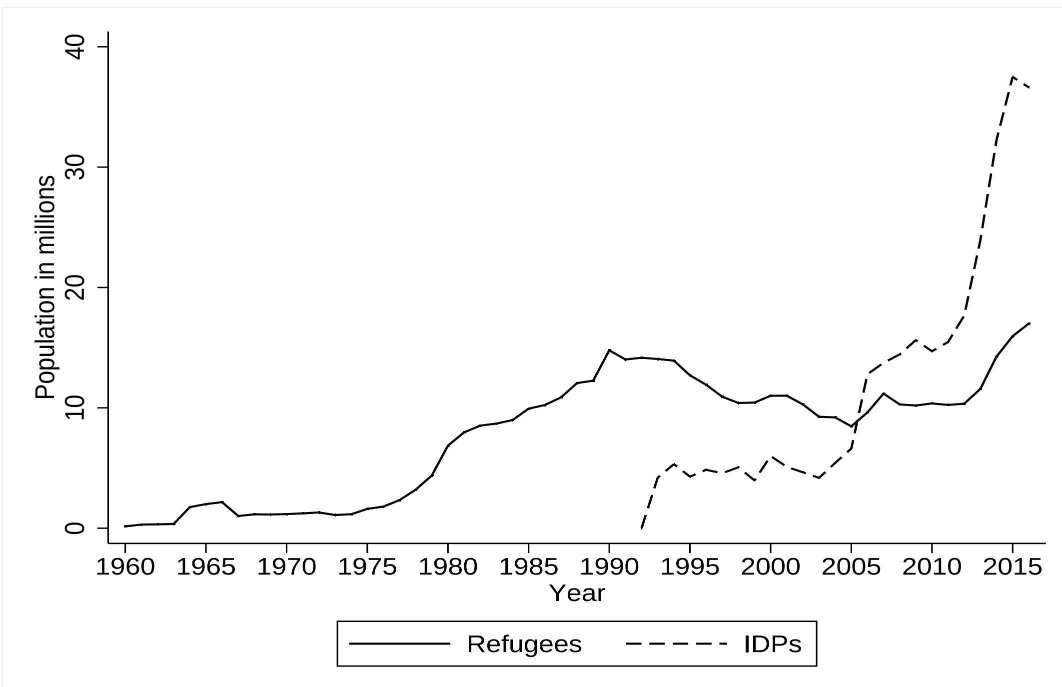

Figure 2.11: Trends in the aggregate number of refugees and IDPs

Figure 2.12 plots the number of countries listed by the UNHCR as the origin of some refugees.

This number experienced a fairly constant growth until the mid 1980s, grew faster during the late 1980s and 1990s and reached almost all the countries by the early 2000's. The empirical evidence reported below refers to the 1995-2016 period, including 200 countries and territories listed in Table 2.1.

Displacement counts prior to 1995 are used to compute the pre-sample mean. The data includes only IDPs uprooted by conflict and human rights violations. Thus, the data does not include displacement caused by natural disasters.[2]

Refugee data are dyadic time series, i.e. each refugee count is associated with an origin country, a destination country and a year. IDP counts are longitudinal data, each involving a country and year. We dropped all refugee counts whose origin were unknown and replaced missing refugee or IDP counts by zero counts.[3]

Next, we aggregated the refugee counts adding all destinations to obtain an unbalanced longitudinal origins dataset.

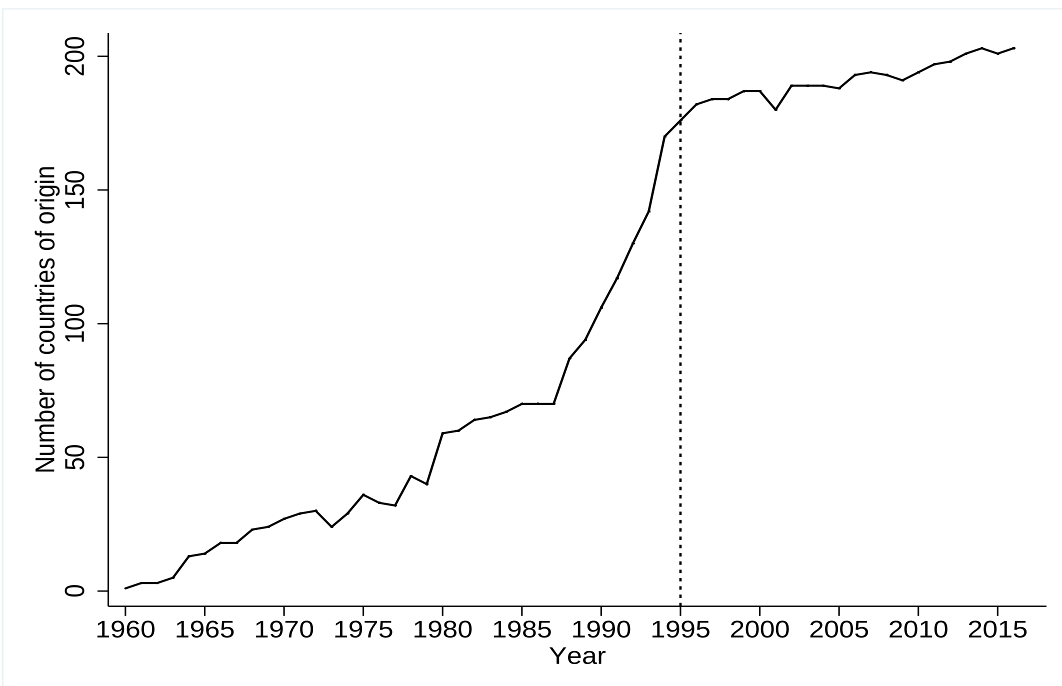

Figure 2.12: Number of countries listed as origin of refugees in the UNHCR population statistics. The vertical line divides the pre-sample (1960-1994) and sample periods (1995-2016).

[2]For a full report on internal displacement see Internal Displacement Monitoring Centre (2018).
[3]Marbach (2018) suggests a method to impute unknown origin refugee counts.

Table 2.1: List of Countries

Countries and territories listed in the UNHCR population statistics included in the analysis:
Afghanistan, Albania, Algeria, Andorra, Angola, Antigua and Barbuda, Argentina, Armenia, Aruba, Australia, Austria, Azerbaijan, Bahamas, Bahrain, Bangladesh, Barbados, Belarus, Belgium, Belize, Benin, Bermuda, Bhutan, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, British Virgin Islands, Brunei, Bulgaria, Burkina Faso, Burundi, Cabo Verde, Cambodia, Cameroon, Canada, Cayman Islands, Central African Republic, Chad, Chile, China, Colombia, Comoros, Congo, Costa Rica, Croatia, Cuba, Cyprus, Czech, North Korea, Denmark, Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Estonia, Ethiopia, Fiji, Finland, France, French Polynesia, Gabon, Gambia, Georgia, Germany, Ghana, Gibraltar, Greece, Grenada, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, Hong Kong, Hungary, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Ivory Coast, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kiribati, South Korea, Kuwait, Kyrgyzstan, Laos, Latvia, Lebanon, Lesotho, Liberia, Libya, Lithuania, Luxembourg, Macao, Macedonia, Madagascar, Malawi, Malaysia, Maldives, Mali, Malta, Marshall Islands, Mauritania, Mauritius, Mexico, Micronesia, Moldova, Mongolia, Morocco, Mozambique, Myanmar, Namibia, Nauru, Nepal, Netherlands, New Caledonia, New Zealand, Nicaragua, Niger, Nigeria, Norway, Oman, Pakistan, Palau, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Puerto Rico, Qatar, Romania, Russia (Soviet Union), Rwanda, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Samoa, San Marino, Sao Tome and Principe, Saudi Arabia, Senegal, Serbia, Seychelles, Sierra Leone, Singapore, Slovakia, Slovenia, Solomon Islands, Somalia, South Africa, Spain, Sri Lanka, Sudan, Suriname, Swaziland, Sweden, Switzerland, Syria, Tajikistan, Tanzania, Thailand, Timor-Leste, Togo, Tonga, Trinidad and Tobago, Tunisia, Turkey, Turkmenistan, Turks and Caicos Islands, Tuvalu, Uganda, Ukraine, United Arab Emirates, United Kingdom, United States of America, Uruguay, Uzbekistan, Vanuatu, Venezuela, Vietnam, Yemen, Zaire, Zambia, Zimbabwe.
Countries and territories listed in the UNHCR population statistics not included in the analysis:
American Samoa, Anguilla, Bonaire, Cook Islands, Curaçao, French Guiana, Guadeloupe, Guam, Liechtenstein, Martinique, Monaco, Montenegro, Montserrat, Niue, Norfolk Island, Palestine, Saint Maarten (Dutch part), Saint Pierre et Miquelon, Svalbard and Jan Mayen, South Sudan, The Holy See, Tibetan, Wallis and Futuna Islands, Western Sahara.

In addition to IDPs and refugee counts, our data set includes the following covariates. Population data from the World Bank, which is used as a measure of exposure, that is, we measure the number of refugees, IDPs and total displacement as a fraction of the origin country's population. Area and the average Riley et al. (1999) index of terrain ruggedness are from Nunn and Puga (2012). Distances between each country's most populated cities are from CEPII. For the empirical analysis, a country's proximity, P_i, is measured as an exponential average of distances from origin country i to all the other N destination countries

$$P_i = \sum_{j=1}^{N} e^{-d_{ij}}$$

where d_{ij} represents the distance between country i and country j, measured as a fraction of the largest distance.[4] Conflict intensity data, from the UCDP-PRIO conflict dataset, are measured as a categorical variable with three categories corresponding to no conflict, low intensity conflict and high intensity conflict.

In addition to these covariates, we also include other control variables regarding which the model does not contribute anything, but have previously been included in empirical models of displacement. First, a large share of forced displacement does not have conflict as the root cause, instead the cause forcing migration is the lack of civil liberties or political rights.

Therefore, we include categorical measures of civil liberties and political rights from Freedom House. Second, it is often argued that refugee counts hide economic migrants, therefore, refugee counts should be negatively related with economic conditions in the source country. Thus, we include GDP per capita from the World Bank in the empirical specification.

Third, another forced migration correlate reported in the literature is colonial relationship. Thus, we include the number of countries with which a source country had a colonial relationship (either as a colony or colonizer). Fourth, in the same vein, we also control for the number of countries with which a source country shares language/ethnicity group in common with at least 9% of the population.

The empirical evidence is complemented by including in the analysis other spatial covariates that we did not account for in the theoretical model but might have a bearing in forced displacement determination. A potentially relevant spatial variable is whether a country's area

[4]Notice that the function $e^{-d_{ij}}$ satisfies the assumptions made about function $\alpha(d_{ij})$.

includes a large desert.

Our model assumes population is uniformly distributed in a country's territory. However, the presence of a desert invalidates our assumption. To account for this issue, we include an indicator of whether more than 20% of a country's area is a desert.

Another spatial feature that is not incorporated in our model is whether a country is landlocked or not. Arguably, a landlocked country, surrounded by other countries, must be easier to escape from than an island country. Therefore, we also include in the analysis an indicator of whether a country is landlocked or not.

2.4.3 Empirics

In this section we report empirical evidence using data from 1995 to 2016. Table 2.2 reports exponential regression estimates for refugee counts. Column (1) reports Pooled-GMM estimates and includes only the spatial variables and conflict intensity, the only variables accounted for in the theoretical model. As predicted by the model, area has a negative and significant effect and proximity a positive and significant one. Ruggedness has a positive and not statistically significant coefficient, while the model predicts a negative partial effect. In accordance with the predictions of the model, conflict intensity has a positive effect on external displacement: high intensity conflicts have a positive and significant effect on refugee counts, while low intensity ones do not affect refugee counts significantly. Next we focus on the PSM-GMM estimates reported in columns (2), (3) and (4). These exponential regressions include the (log of) the mean value of the endogenous variable (the number of refugees as a fraction of population) during the pre-sample period, in our case, 1960-1994. Therefore, the sample is restricted to 168 countries with positive pre-sample mean, so that the logarithm can be computed. Adding the pre-sample mean to account for unobserved heterogeneity in column (2) improves the goodness of fit considerably, as measured by the pseudo-R^2. In addition, the coefficient on ruggedness turns out to be negative, as predicted by the model, although it is significant only at ten percent. Furthermore, the low intensity conflict indicator also becomes significant at the ten per cent level. Finally, the pre-sample mean is positive and significant, thus indicating its relevance in accounting for heterogeneity. Column (3) in Table 2.2 includes other determinants of refugee counts found relevant in the literature. Civil liberties are measured on a scale of 1 to 7, with 1 representing the highest standards of civil liberties and 7 the lowest. We incorporate this variable using 6 dummy variables for levels 2 to 7, leaving level 1 (the highest standards) as the reference group. It turns out that as we move to lower civil liberties standards (from level 2 to level 7) the estimated coefficients are larger. In other words, refugee counts are uniformly increasing as civil liberties standards deteriorate. On the other hand, regarding the effect of economic conditions at the source countries, we find that higher levels of real GDP per capita in origin countries are significantly associated with lower refugee counts. However, neither the common linguistic-ethnicity proxy nor the colonial-relation measure turn out to be significant. Overall, including civil liberties and GDP per capita improves the goodness of fit and increases the significance of the spatial and conflict intensity variables. Column (4) accounts for other spatial variables not considered in the theoretical model that might be relevant for refugee flows in practice. Neither the desert nor the landlocked indicators are significant, while the results for the other covariates remain almost invariant quantitatively and in terms of significance. Replacing the civil liberties with political rights yields very similar results, which are not reported and available as an online appendix.

Table 2.2: Exponential regressions: refugees

	(1)	(2)	(3)	(4)
(Log) Area	-0.4904***	-0.3153***	-0.3405***	-0.3408***
	(0.0814)	(0.0943)	(0.0688)	(0.0662)
(Log) Ruggedness	0.3422	-0.2569*	-0.3407**	-0.3784**
	(0.2526)	(0.1428)	(0.1544)	(0.1883)
(Log) Proximity	0.1169***	0.1017***	0.0823**	0.0875**
	(0.0419)	(0.0393)	(0.0400)	(0.0421)
Low intensity conflict	0.5232	0.4658*	0.4262**	0.4034*
	(0.3715)	(0.2708)	(0.2145)	(0.2174)
High intensity conflict	2.7842***	1.4917***	0.8885***	0.8442***
	(0.4277)	(0.3826)	(0.2086)	(0.2133)
Civil Liberties level 2			2.0607***	2.0685***
			(0.4322)	(0.4324)
Civil Liberties level 3			2.0913***	2.1427***
			(0.4543)	(0.4756)
Civil Liberties level 4			2.7370***	2.7539***
			(0.5185)	(0.5443)
Civil Liberties level 5			3.0028***	3.0490***
			(0.5868)	(0.6172)
Civil Liberties level 6			3.3458***	3.4084***
			(0.5667)	(0.6008)
Civil Liberties level 7			3.9158***	3.9849***
			(0.6600)	(0.7246)
(Log) GDP per capita			-0.3777**	-0.3915**
			(0.1777)	(0.1855)
(Log) # common ethnicity			0.1095	0.0983
			(0.1125)	(0.1093)
(Log) # colonial relations			0.1018	0.1094
			(0.3295)	(0.3329)
% desert > 20 %				-0.5911
				(0.6506)
Landlocked				0.0382
				(0.3344)
Pre-Sample Mean		0.4631***	0.4851***	0.4786***
		(0.0908)	(0.0485)	(0.0497)
Constant	-7.8443***	-3.9968***	-2.6189	-2.6316
	(0.8706)	(1.0009)	(2.4906)	(2.6564)
Pseudo - R^2	0.4695	0.7596	0.8258	0.8269
Number of observations	4,395	3,691	3,455	3,455
Number of countries	200	168	161	161

Lagged population used as exposure in all regressions

Regressions include year and continent dummies.

Country-clustered robust standard errors in parentheses.

One, two and three stars stand for 10, 5 and 1 per cent significance levels.

Table 2.3 reports the same exponential regressions as in Table 2.2 with two differences. First, the dependent variable is now IDP counts. Second, Data on IDPs are available from 1993 to 2016. Thus, the pre-sample period includes only IDP counts for 1993 and 1994, which leaves just two IDP counts to compute the pre-sample mean. However, under the assumptions of our model, at the early stages of displacement, all people displaced are IDPs and at later stages some become refugees. A positive refugee count is therefore evidence of earlier internal displacement. This reasoning suggests that refugee counts must be preceded by IDP counts. Accordingly, the pre-sample mean used in columns (2) to (4) corresponds to the pre-sample mean of total displacement, that is, the sum of refugees and IDPs. The results reported in Table 2.3 are very similar to those reported in Table 2.2, both quantitatively and in terms of significance. The only differences are that civil liberties and GDP per capita are somewhat less significant for IDP regressions than for refugee ones. The second level of civil liberties, below the highest standard, is not significantly different from the reference group (civil liberties level 1, the highest standard). However, civil liberties level 7 (the lowest standards) is highly significant and has a quantitatively similar coefficient estimate to the refugees regression. IDP counts are not strictly uniformly decreasing in civil liberties, but almost. In accordance with the model's prediction, the coefficients on the conflict intensity indicators are positive and statistically significant. Coefficient estimates on area, ruggedness and proximity are significant and have the same signs as in the refugee regression. Appendix 2.7 shows that the model predicts that the partial derivatives of IDPs with respect to the spatial variables might be either positive or negative depending on spatial variables and conflict intensity values. Table 2.3 indicates that area has a negative partial effect, which is consistent with the theory provided proximity is not very high. Ruggedness affects internal displacement negatively and significantly, which is in line with the model only for low values of proximity. Proximity has a positive coefficient on the IDPs regression, as predicted by the model for high shock intensity. While this evidence fits the predictions of the model, strictly speaking, we cannot validate nor refute the model based on the results from the IDPs regressions.

Table 2.3: Exponential regressions: IDPs

	(1)	(2)	(3)	(4)
(Log) Area	-0.6700***	-0.5265***	-0.4660***	-0.4864***
	(0.1167)	(0.1320)	(0.1319)	(0.1264)
(Log) Ruggedness	-0.1353	-0.5013***	-0.6382***	-0.6126***
	(0.1772)	(0.1877)	(0.2069)	(0.2189)
(Log) Proximity	0.2325***	0.2056***	0.2259***	0.2317***
	(0.0625)	(0.0581)	(0.0784)	(0.0775)
Low intensity conflict	2.2796***	2.1725***	1.6122***	1.6423***
	(0.5084)	(0.5363)	(0.4776)	(0.4684)
High intensity conflict	3.7109***	3.1088***	2.0410***	2.0134***
	(0.5821)	(0.6249)	(0.6345)	(0.6306)
Civil Liberties level 2			0.1799	-0.0497
			(1.5346)	(1.5198)
Civil Liberties level 3			1.9730*	1.6507
			(1.1813)	(1.1878)
Civil Liberties level 4			2.5737**	2.2679*
			(1.2014)	(1.2030)
Civil Liberties level 5			2.4682**	2.2481*
			(1.2539)	(1.2745)
Civil Liberties level 6			2.5747*	2.2566*
			(1.3319)	(1.3183)
Civil Liberties level 7			3.4573***	3.0998**
			(1.3349)	(1.3512)
(Log) GDP per capita			-0.5125*	-0.7030**
			(0.3007)	(0.3119)
(Log) # common ethnicity			0.1874	0.1886
			(0.1719)	(0.1841)
(Log) # colonial relations			0.1733	0.2203
			(0.2859)	(0.2834)
% desert > 20 %				-0.5571
				(0.8136)
Landlocked				-0.8236
				(0.6076)
Pre-Sample Mean		0.2708***	0.2992***	0.3232***
		(0.0604)	(0.0603)	(0.0573)
Constant	-10.8868***	-8.3968***	-4.0450	-1.9149
	(1.1515)	(1.2356)	(4.0938)	(4.0484)
Pseudo - R^2	0.5702	0.6704	0.6988	0.7108
Number of observations	4,395	3,691	3,455	3,455
Number of countries	200	168	161	161

Lagged population used as exposure in all regressions which include year and continent dummies. Country-clustered robust standard errors in parentheses. One, two and three stars stand for 10, 5 and 1 per cent significance levels.

Table 2.4 reports exponential regression results for total displacement which are similar to those found for internal and external displacement. In particular, the spatial variables, area, ruggedness and proximity are significant and have the same signs as in the previous tables. Area and ruggedness have a negative effect on total displacement, as suggested by the model. However, the model predicts proximity should not have any bearing in determining total displacement, while we find a positive and significant effect. A feasible explanation of this outcome is that total displacement and, in particular, refugee counts record not only conflict induced displacement, as assumed in the model, but also those who flee their country for political, religious, ethnic, gender or other reasons, and also might include some economic migrants. All these people might be more likely to flee their country of origin for higher proximity levels. Results not reported but available upon request show that the findings reported in Tables 2.2-2.4 remain very much the same when the civil liberties indicator is replaced with a political rights indicator. Overall, our reading of the results is that they fundamentally agree with the predictions of the theoretical model. First, all the evidence uses population as exposure, as indicated by the model, and the specification choice results in good fit for this type of data. Second, the spatial variables have the predicted coefficient signs in the regressions, with the only exception of proximity in the total displacement regressions. Third, all displacement counts (refugees, IDPs and total) are uniformly increasing in conflict intensity. Finally, the other determinants of forced migration included in the regressions but not accounted for in the theoretical model have coefficient signs which accord with previous findings in the literature.

Table 2.4: Exponential regressions: total displacement

	(1)	(2)	(3)	(4)
(Log) Area	-0.5852***	-0.4046***	-0.3970***	-0.4105***
	(0.0975)	(0.1059)	(0.0852)	(0.0801)
(Log) Ruggedness	0.1133	-0.3564**	-0.4557***	-0.4377**
	(0.2048)	(0.1489)	(0.1674)	(0.1812)
(Log) Proximity	0.1636***	0.1403***	0.1339***	0.1434***
	(0.0507)	(0.0403)	(0.0430)	(0.0425)
Low intensity conflict	1.4362***	1.3853***	1.0883***	1.0843***
	(0.4243)	(0.4216)	(0.3233)	(0.3141)
High intensity conflict	3.1726***	2.3605***	1.4707***	1.4320***
	(0.3818)	(0.4185)	(0.3843)	(0.3683)
Civil Liberties level 2			0.9272	0.8299
			(1.0391)	(1.0420)
Civil Liberties level 3			2.0201**	1.8588*
			(0.9441)	(0.9530)
Civil Liberties level 4			2.7725***	2.6221***
			(0.9173)	(0.9242)
Civil Liberties level 5			2.8663***	2.7531***
			(0.9622)	(0.9713)
Civil Liberties level 6			3.0599***	2.9244***
			(0.9839)	(0.9812)
Civil Liberties level 7			3.9046***	3.7649***
			(1.0253)	(1.0367)
(Log) GDP per capita			-0.4234**	-0.5496***
			(0.1991)	(0.2033)
(Log) # common ethnicity			0.1803	0.1816
			(0.1274)	(0.1290)
(Log) # colonial relations			0.0263	0.0714
			(0.2165)	(0.2150)
% desert > 20 %				-0.5094
				(0.5900)
Landlocked				-0.4412
				(0.3352)
Pre-Sample Mean		0.3380***	0.3555***	0.3671***
		(0.0616)	(0.0449)	(0.0479)
Constant	-8.4598***	-5.5005***	-3.2830	-1.8899
	(0.9494)	(0.9710)	(2.7763)	(2.7361)
Pseudo - R^2	0.5189	0.7067	0.7616	0.7676
Number of observations	4,395	3,691	3,455	3,455
Number of countries	200	168	161	161

Lagged population used as exposure in all regressions which include year and continent dummies. Country-clustered robust standard errors in parentheses. One, two and three stars stand for 10, 5 and 1 per cent significance levels.

2.5 Conclusions

This chapter proposes a spatial model to analyze internal and external forced migration, a topic of increasing relevance nowadays. The bare bones model worked out in this chapter does not pretend to develop a new theory about how internal and external forced displacement are generated. On the contrary, it recognizes that previous studies have shown that armed conflict and lack of political rights or civil liberties cause people displacement. This is to say that the root causes of forced displacement are well known. The contribution of this chapter is to analyze the role of spatial factors in the relationship between the root causes (armed conflict and lack of liberties) and forced displacement. In the causal inference jargon, these spatial factors are called effect modifiers. The model predicts how some spatial factors are related with internal, external and total displacement. In particular, we focus on country size (measured by area), how abrupt the geography of a country is (measured by its ruggedness) and the degree of geographical proximity of the country with respect to the rest of the world. The model also predicts how conflict intensity ought to be related to internal displacement and forced migration. In addition, the model suggests that population should be used as a measure of exposure, and that is how it enters the empirical analysis. We test these predictions against real data. The evidence reported is broadly consistent with the predictions of the model. We chose to develop a barebones model intentionally, hoping that its simplicity would seduce other scholars. There are a number of feasible extensions to this work. First, the model can be developed using other spatial frameworks by assuming that the origin country is, say a circle, or other two- or higher-dimensional figures. The model could also be solved under different assumptions about how shocks affect people, induced displacement, costs of displacement, cost of settlement and other issues we have not taken into account in our simplified model. The model could also be used to model displacement provoked by natural disasters and the predictions of such a model could be tested using data from displacement induced by natural disasters. Finally, although our model focuses on the country of origin, the analysis could be extended to destination countries, or even to dyadic flows.

2.6 Appendix B: Data sources of Chapter 2

The data set compiled is a blend of the following data sets:

1. Persons of concern Dataset. United Nations High Commissioner for the Refugees (UNHCR). Time frame: 1951-2016. Variables used: Refugees, IDPs.

   Availability: `http://popstats.unhcr.org/en/persons_of_concern`

2. Armed Conflict Dataset. Uppsala Conflict Data Program (UCDP) Uppsala University / Peace Research Institute Oslo (PRIO). Time frame: 1946-2015. Variables used: Location. The name(s) of the country/countries whose government(s) have a primary claim to the issue in dispute. Year of observation. The date when the conflict activity reached 25 battle-related deaths in a year. The date when conflict activity ended. The intensity level of the armed conflict.

   Availability: `https://www.prio.org/Data/Armed-Conflict/UCDP-PRIO/`

3. Freedom in the World Dataset. Freedom House. Time frame: 1972-2015. Variables used: Civil Liberties and Political Rights.

   Availability: `https://freedomhouse.org`

4. Geographical and Distance dataset. CEPII, SciencesPo Department of Economics. Dyadic data set. Variables used: Common language/ethnicity indicator. Colonial relationship indicator. Simple distance (most populated cities, km).

   Availability: `http://econ.sciences-po.fr/staff/thierry-mayer`

5. World Development Indicators. The World Bank. Population. GDP per capita, PPP (constant 2011 international $).

   Availability: `http://databank.worldbank.org/data/home.aspx`

6. Geographical variables. Land area. Riley's Index of Terrain Ruggedness. From Nunn and Puga (2012).

   Availability: `https://diegopuga.org/data/rugged/`

2.7 Appendix C: Expected values and derivatives of Chapter 2

To integrate out the location we proceed in two regions defined by the intersection of the two lines in Figure 2.2. First we integrate 2.2 for values of the shock in the interval $[0,\frac{1}{2}]$ and then in the interval $\left(\frac{1}{2},1\right]$

$$E_{l_s}[D|s] = \begin{cases} \int_s^{1-s} 2s\mathrm{d}l_s + \int_0^s (l_s+s)\mathrm{d}l_s + \int_{1-s}^1 (s+1-l_s)\mathrm{d}l_s = 2s-s^2 \quad if \quad s\in[0,\frac{1}{2}] \\ \\ \int_0^{1-s} 2s\mathrm{d}l_s + \int_s^1 (s+1-l_s)\mathrm{d}l_s + \int_{1-s}^s dl_s = 2s-s^2 \quad if \quad s\in\left(\frac{1}{2},1\right] \end{cases}$$

Next we integrate 2.3 in four regions defined by the intersections of the four lines in Figure 2.6. For values of the shock in the intervals $\left[0,\frac{1}{4}\right]$, $\left(\frac{1}{4},\frac{1}{3}\right]$, $\left(\frac{1}{3},\frac{1}{2}\right]$ and $\left(\frac{1}{2},1\right]$

$$E_{l_s}[R|s] = \begin{cases} \int_0^s (l_s)\mathrm{d}l_s + \int_s^{2s}(2s-l_s)\mathrm{d}l_s \\ \quad + \int_{1-2s}^{1-s}(l_s+2s-1)\mathrm{d}l_s + \int_{1-s}^1 (1-l_s)\mathrm{d}l_s \quad if \quad s\in[0,\frac{1}{4}] \\ \\ \int_0^s (l_s)\mathrm{d}l_s + \int_s^{1-2s}(2s-l_s)\mathrm{d}l_s + \int_{1-2s}^{2s}(4s-1)\mathrm{d}l_s \\ \quad + \int_{2s}^{1-s}(l_s+2s-1)\mathrm{d}l_s + \int_{1-s}^1 (1-l_s)\mathrm{d}l_s \quad if \quad s\in(\frac{1}{4},\frac{1}{3}] \\ \\ \int_0^{1-2s}(l_s)\mathrm{d}l_s + \int_{1-2s}^s (2s+2l_s-1)\mathrm{d}l_s + \int_s^{1-s}(4s-1)\mathrm{d}l_s \\ \quad + \int_{1-s}^{2s}(2s+1-2l_s)\mathrm{d}l_s + \int_{2s}^1 (1-l_s)\mathrm{d}l_s \quad if \quad s\in(\frac{1}{3},\frac{1}{2}] \\ \\ \int_0^{1-s}(2l_s+2s-1)\mathrm{d}l_s + \int_{1-s}^s dl_s + \int_s^1 (2s+1-2l_s)\mathrm{d}l_s \quad if \quad s\in(\frac{1}{2},1] \end{cases}$$

which results in the much simpler expression

$$E_{l_s}[R|s] = \begin{cases} 2s^2 & if \quad s \in [0, \frac{1}{2}] \\ \\ 4s - 2s^2 - 1 & if \quad s \in (\frac{1}{2}, 1] \end{cases}$$

The fraction of the population that remain as IDPs is found by subtraction

$$E_{l_s}[I|s] = E[D] - E[R] = \begin{cases} 2s - 3s^2 & if \quad s \in [0, \frac{1}{2}] \\ \\ s^2 - 2s + 1 & if \quad s \in (\frac{1}{2}, 1] \end{cases}$$

Next consider the model of Section 2.3.1. Integrating equation 2.7

$$E_{l_s}[D|s,a,r,d_{AB},d_{BC}] = \begin{cases} [\int_0^a (2\gamma s)/\mathrm{a}\,\mathrm{d}l_s - \int_0^{\gamma s} (\gamma s - l_s)/\mathrm{a}\,\mathrm{d}l_s & \\ \quad - \int_{a-\gamma s}^a (l_s + \gamma s - a)/\mathrm{a}\,\mathrm{d}l_s]/a & if \quad s \in [0, a/2] \\ \\ [\int_0^a (2\gamma s)/a\,dl_s - \int_0^{a-\gamma s} (\gamma s - l_s)/\mathrm{a}\,\mathrm{d}l_s & \\ \quad - \int_{\gamma s}^a (l_s + \gamma s - a)/a\,dl_s & \\ - \int_{a-\gamma s}^{\gamma s} [(\gamma s - l_s) + (l_s + \gamma s - a)]/\mathrm{a}\,\mathrm{d}l_s]/a & if \quad s \in (a/2, a] \end{cases}$$

$$= (2\gamma as - \gamma^2 s^2)/a^2 \quad \in [0, a]$$

Similarly integrating 2.8 we obtain

$$
E_{l_s}[R|s,a,r,d_{AB},d_{BC}] =
\begin{cases}
\begin{array}{c}
[\int_0^{\gamma s} \alpha_{A,B}(l_s)/\mathrm{a}\,\mathrm{d}l_s + \int_{\gamma s}^{2\gamma s} \alpha_{A,B}(2\gamma s - l_s)/\mathrm{a}\,\mathrm{d}l_s \\
+ \int_{a-2\gamma s}^{a-\gamma s} \alpha_{B,C}(l_s + 2\gamma s - a)/\mathrm{a}\,\mathrm{d}l_s \\
+ \int_{a-\gamma s}^{a} \alpha_{B,C}(a - l_s)/\mathrm{a}\,\mathrm{d}l_s]/a
\end{array} & if \quad s \in [0, a/4] \\
\begin{array}{c}
[\int_0^{\gamma s} \alpha_{A,B}(l_s)/\mathrm{a}\,\mathrm{d}l_s + \int_{\gamma s}^{a-2\gamma s} \alpha_{A,B}(2\gamma s - l_s)/\mathrm{a}\,\mathrm{d}l_s \\
+ \int_{a-2\gamma s}^{2\gamma s} [\alpha_{A,B}(2\gamma s - l_s) + \alpha_{B,C}(l_s + 2\gamma s - a)]/\mathrm{a}\,\mathrm{d}l_s \\
+ \int_{2\gamma s}^{a-\gamma s} \alpha_{B,C}(l_s + 2\gamma s - a)/\mathrm{a}\,\mathrm{d}l_s \\
+ \int_{a-\gamma s}^{a} \alpha_{B,C}(a - l_s)/a\,dl_s]/a
\end{array} & if \quad s \in (a/4, a/3] \\
\begin{array}{c}
[\int_0^{a-2\gamma s} \alpha_{A,B}(l_s)/\mathrm{a}\,\mathrm{d}l_s \\
+ \int_{a-2\gamma s}^{\gamma s} \alpha_{A,B}(l_s)/a + \alpha_{B,C}(l_s + 2\gamma s - a)/\mathrm{a}\,\mathrm{d}l_s \\
+ \int_{\gamma s}^{a-\gamma s} \alpha_{A,B}(2\gamma s - l_s)/a \\
+ \alpha_{B,C}(l_s + 2\gamma s - a)/\mathrm{a}\,\mathrm{d}l_s \\
+ \int_{a-\gamma s}^{2\gamma s} \alpha_{B,C}(2\gamma s + a - 2l_s)/\mathrm{a}\,\mathrm{d}l_s \\
+ \int_{2\gamma s}^{a} 1/\alpha_{B,C}(a - l_s)/\mathrm{a}\,\mathrm{d}l_s]/a
\end{array} & if \quad s \in (a/3, a/2] \\
\begin{array}{c}
[\int_0^{a-\gamma s} \alpha_{A,B}\, l_s/a + \alpha_{B,C}(l_s + 2\gamma s - a)/\mathrm{a}\,\mathrm{d}l_s \\
+ \int_{a-\gamma s}^{\gamma s} \alpha_{A,B}\, l_s/a \\
+ \alpha_{B,C}(a - l_s)/a\,dl_s \\
+ \int_{\gamma s}^{a} \alpha_{A,B}(2\gamma s - l_s)/a + \alpha_{\mathrm{B,C}}(\mathrm{a} - \mathrm{l_s})/\mathrm{a}\,\mathrm{d}l_s]/a
\end{array} & if \quad s \in (a/2, a]
\end{cases}
$$

$$= \begin{cases} [\gamma^2 s^2 (\alpha_{A,B} + \alpha_{B,C})]/a^2 & if \quad s \in [0, a/4] \\ [\gamma^2 s^2 (\alpha_{A,B} + \alpha_{B,C})]/a^2 & if \quad s \in (a/4, a/3] \\ [\gamma^2 s^2 (\alpha_{A,B} + \alpha_{B,C})]/a^2 & if \quad s \in (a/3, a/2] \\ [(2\gamma as - \gamma^2 s^2 - a^2/2)(\alpha_{A,B} + \alpha_{B,C})]/a^2 & if \quad s \in (a/2, a] \end{cases}$$

In particular, we are interested in the derivatives of (2.4), (2.5) and (2.6) with respect to area, ruggedness and proximity. These are as follows:

$$\frac{\partial E_{l_s}[D|s,a,\gamma(r),P_B]}{\partial a} = \frac{2\gamma s}{a^3}(\gamma s - a) \leq 0$$
$$\frac{\partial E_{l_s}[D|s,a,\gamma(r),P_B]}{\partial r} = \frac{2\gamma s}{a^2}(a - \gamma s)\gamma'(r) \leq 0,$$
$$\frac{\partial E_{l_s}[D|s,a,\gamma(r),P_B]}{\partial P_B} = 0.$$

Ruggedness and country size reduce the fraction of the population displaced while proximity does not affect total displacement. The fraction of the population that become refugees depends on these magnitudes in a way that depends on whether the shock is below or above $a/2$. In particular, we have that the derivatives of the fraction of the population that become refugees are

$$\frac{\partial E_{l_s}[R|s,a,\gamma(r),P_B]}{\partial a} = \begin{cases} -\frac{2\gamma^2 s^2 P_B}{a^3} \leq 0 & if \quad s \in [0, \frac{a}{2}] \\ \frac{2\gamma s(\gamma s - a)P_B}{a^3} < 0 & if \quad s \in (\frac{a}{2}, a] \end{cases}$$

$$\frac{\partial E_{l_s}[R|s,a,\gamma(r),P_B]}{\partial r} = \begin{cases} \frac{2\gamma'(r)\gamma s P_B}{a^2} \leq 0 & if \quad s \in \left[0, \frac{a}{2}\right] \\ \\ \frac{2s\gamma'(r)(a-\gamma s)P_B}{a^2} \leq 0 & if \quad s \in \left(\frac{a}{2}, a\right] \end{cases}$$

$$\frac{\partial E_{l_s}[R|s,a,\gamma(r),P_B]}{\partial P_B} = \begin{cases} \left(\frac{\gamma s}{a}\right)^2 \geq 0 & if \quad s \in [0, \frac{a}{2}] \\ \\ \frac{1}{2} - \frac{(\gamma s - a)^2}{a^2} > 0 & if \quad s \in (\frac{a}{2}, a] \end{cases}$$

Therefore, country size and ruggedness affect negatively and proximity positively the fraction of the population that become refugees. Finally, the fraction of the population that become IDPs depends on country size, ruggedness and proximity as follows

$$\frac{\partial E_{l_s}[I|s,a,\gamma(r),P_B]}{\partial a} = \begin{cases} \frac{2\gamma s(\gamma s(1+P_B)-a)}{a^3} & if \quad s \in [0, \frac{a}{2}] \\ \\ \frac{2\gamma s(\gamma s-a)(1-P_B)}{a^3} & if \quad s \in (\frac{a}{2}, a], \end{cases}$$

$$\frac{\partial E_{l_s}[I|s,a,\gamma(r),P_B]}{\partial r} = \begin{cases} -\frac{2\gamma'(r)s(\gamma s(1+P_B)-a)}{a^2} & if \quad s \in (0, \frac{a}{2}] \\ \\ -\frac{2\gamma'(r)s(1-P_B)(\gamma s-a)}{a^2} & if \quad s \in (\frac{a}{2}, a], \end{cases}$$

$$\frac{\partial E_{l_s}[I|s,a,\gamma(r),P_B]}{\partial P_B} = \begin{cases} -\left(\frac{\gamma s}{a}\right)^2 & if \quad s \in (0, \frac{a}{2}] \\ \\ \frac{1}{2} - \frac{\gamma s(2a-\gamma s)}{a^2} & if \quad s \in (\frac{a}{2}, a]. \end{cases}$$

While the signs of the derivatives of total displacement and refugees are uniform on s, this is no longer the case for the signs of these derivatives of internal displacement

	$0 \leq P_B \leq 1$	$1 < P_B \leq 2$	
		$0 < \gamma \leq \frac{2}{3}$	$\frac{2}{3} < \gamma \leq 1$
$0 \leq s \leq \frac{a}{2}$	$\frac{\partial E_{l_s}[I\|s,a,\gamma(r),P_B]}{\partial a} \leq 0$	$\frac{\partial E_{l_s}[I\|s,a,\gamma(r),P_B]}{\partial a} \leq 0$	$\frac{\partial E_{l_s}[I\|s,a,\gamma(r),P_B]}{\partial a} > 0$
$\frac{a}{2} < s \leq a$	$\frac{\partial E_{l_s}[I\|s,a,\gamma(r),P_B]}{\partial a} \leq 0$	$\frac{\partial E_{l_s}[I\|s,a,\gamma(r),P_B]}{\partial a} > 0$	

	$0 \leq P_B \leq 1$	$1 < P_B \leq 2$	
		$0 < \gamma \leq \frac{2}{3}$	$\frac{2}{3} < \gamma \leq 1$
$0 \leq s \leq \frac{a}{2}$	$\frac{\partial E_{l_s}[I\|s,a,\gamma(r),P_B]}{\partial r} \leq 0$	$\frac{\partial E_{l_s}[I\|s,a,\gamma(r),P_B]}{\partial r} \leq 0$	$\frac{\partial E_{l_s}[I\|s,a,\gamma(r),P_B]}{\partial r} > 0$
$\frac{a}{2} < s \leq a$	$\frac{\partial E_{l_s}[I\|s,a,\gamma(r),P_B]}{\partial r} \leq 0$	$\frac{\partial E_{l_s}[I\|s,a,\gamma(r),P_B]}{\partial r} > 0$	

	$0 < \gamma < \frac{(2-\sqrt{2})a}{2s}$	$\frac{(2-\sqrt{2})a}{2s} \leq \gamma \leq 1$
$0 \leq s \leq \frac{a}{2}$	$\frac{\partial E_{l_s}[I\|s,a,\gamma(r),P_B]}{\partial P_B} \leq 0$	
$\frac{a}{2} < s \leq a$	$\frac{\partial E_{l_s}[I\|s,a,\gamma(r),P_B]}{\partial P_B} > 0$	$\frac{\partial E_{l_s}[I\|s,a,\gamma(r),P_B]}{\partial P_B} \leq 0$

Chapter 3

Gender bias in refugee destination flows

Abstract

Men outnumber women in refugee flows. This chapter studies this difference in the number of male and female refugees, i.e. the gender bias in destination flows. In order to analyze the gender bias, we take into account factors such as civil liberties and per capita GDP of destination countries, country of origin-related variables weighted by distance and measures of women's rights. The evidence suggests that the gender bias is greater (more male refugees than female) in women-friendly destination countries and ones with high civil liberty standards, isolated from poor and low civil liberty standards origin countries.

3.1 Introduction

Refugees abandon their origin countries escaping from armed conflicts or persecution. As argued in Echevarria and Gardeazabal (2016), refugees are likely to head for destination countries where armed conflicts do not occur and civil liberties are respected. In other words, countries without armed conflicts, and with high levels of civil liberties are prone to receiving a larger number of refugees. However, the gender balance of these refugee flows is not clear. This chapter contributes a gender perspective to the forced migration literature. We study the extent to which refugee arrivals in destination countries exhibit a gender bias and its determinants. We define the gender bias in refugee destination flows as the difference in the number of male and female refugees. The situation of female refugees has been studied from many perspectives, but there are not studies regarding the gender profile of refugee destination flows, either on a country-level, or on a global level. This chapter is an attempt to fill this gap in the literature. There are several reasons to believe that the number of male and female refugees might not be balanced. First, women's economic and social rights are fewer than men's in many countries and therefore, women might have more reasons than men to flee from countries where their rights are not respected. Second, being discriminated against, women might find it more difficult to abandon their home country. Third, female refugees might have more difficulty than male refugees in finding a destination country where their freedom is respected and security is not an issue. Among the determinants of the gender bias in refugee flows we consider destination-specific variables such as civil liberties, per capita GDP, the degree of isolation of each destination country with respect to origin countries, and the existence of armed conflicts in destination countries.

Intuition suggests that rich, free and close destination countries with no armed conflicts might host more refugees. However, we have no real guide as to how these factors could generate a greater proportion of male or female refugees in destination countries. To provide greater flexibility to the specification, we also add interactions between per capita GDP and civil liberties.

In Echevarria and Gardeazabal (2016), we saw that both origin -and destination- specific variables play a key role in understanding dyadic refugees counts. However, in this chapter we study the number of refugees hosted by each destination country. We use the distance between countries to aggregate origin-related variables and include them in our estimations. For each destination country, we measure the per capita GDP of all origin countries weighted by distance. Thus, the resulting variable measures how close or far a destination country is from rich or poor origin countries. Similarly, for each destination, we measure the level of civil liberties in all origin countries weighted by distance, to come up with a measure of how far a destination country is from countries with high or low civil liberty standards. In addition, for each destination, we also measure the interaction between GDP and civil liberties in each origin country weighted by distance. Again, intuition suggests that destination countries close to authoritarian and poor origin countries might host a greater proportion of refugees, but it is not clear whether these variables ought to affect the gender bias.

However, all these determinants of refugee counts might affect the number of male and female refugees differently, thus generating a gender bias. Another covariate accounted for in our specification search is the World Bank's Women, Business and Law index in destination countries. This index measures women's rights, such as the freedom of movement or the ability to have access to money. Intuition suggests that a country scoring low in this index should receive less female refugees than other countries. This chapter contributes empirical evidence divided into two set of estimations: we study the gender bias without gender-related variables, on the one hand, and with gender-related variables (the Women, Business and Law index) on the other hand. We use a panel data set of 154 countries covering the 2008-2016 period, for the estimations without gender-related variables, and a panel data set of 137 countries in the period 2009-2016 for the estimations with gender-related variables. Our empirical results suggest that a greater proportion of female refugees go to destination countries where civil liberty is not respected, while the proportion of male refugees is greater in countries with high civil liberty standards.

The gender bias is smaller in destination countries far from rich and free origin countries. On the contrary, more male refugees arrive in destination countries isolated from the poorest origin countries with low civil liberty levels. As regards women's economic or social rights, destination countries close to poor origin countries with a smaller Women, Business and Law index receive a greater proportion of female refugees. This result does not fit the intuition that women should go where their liberties are respected. Previous literature regarding refugee destination flows, and the economic and social rights of female refugees is mentioned in Section 3.2. Section 3.3 describes UNHCR data on refugees. The empirical analysis is shown in Section 3.4, while Section 3.5 sums up the main conclusions.

3.2 Literature review

This study analyzes the gender bias in refugee destinations. The literature can be classified into two areas: refugee destination studies and women's economic and social rights studies.

Refugee destination flows have been studied from many perspectives: Groen (2016) proposes a simulation of refugee movements through a network-based agent based model in order to predict the most likely destination countries for the 2012 Northern Mali Conflict.

Suleimenova et al. (2017) present a generalized simulation development approach (SDA) to predict how refugees are distributed across camps in destination countries, given a particular conflict situation, for the cases of Burundi, Central African Republic and Mali.

Previous literature has mentioned many determinants of refugee flows. Moore and Shellman (2007a) do a global dyadic analysis of the pull-push characteristics of refugee flows by introducing factors which determine why refugees leave origin countries, and why refugees go to specific destination countries. Moore and Shellman (2007a) find that a high level of democracy and per capita GNP attract more refugees to destination countries.

Echevarria and Gardeazabal (2016) propose a gravity model to investigate the determinants of global refugee flows, introducing origin and destination-specific factors, such as the level of civil liberties and per capita GDP of destination countries. In our study, we include the civil liberties index and the per capita GDP of destination countries as factors of the gender bias. We analyze whether these determinants imply a greater or lower gender bias. Neumayer (2004) and Neumayer (2005b) focus their research on Western European countries to determine the factors which attract more refugees to such destinations. Schaeffer (2010) introduces a theoretical model to study whether prospective refugees decide to stay in their origin countries, or leave to specific destination countries.

The Havinga and Böcker (1999) study for Belgium, Netherlands and the UK finds that the main factors behind asylum-seeker's choice of destination country, are ties between the country of origin and the country of asylum as well as the characteristics of destination countries. Echevarria and Gardeazabal (2016) introduce another covariate to study dyadic refugee flows: the distance between origin and destination countries. They conclude that distance is one of the main factors which reduce refugee flows between origin and destination countries. We introduce the isolation of destination countries in our analysis in order to determine its effect on the gender bias.

Armed conflict is one of the most important determinants of refugee flows. Therefore, this is another determinant we include in our estimations of the gender bias, taking into account that armed conflicts in destination countries reduce refugee flows in destination countries. Böhmelt et al. (2019) study the effect of refugee destination flows on the risk of armed conflicts between organized non-state groups, potentially creating a pull effect on such flows.

Milton et al. (2013) analyze the relationship between refugee flows and transnational terrorism in destination countries. They point out the conditions in refugee camps and the treatment host states give to refugees as possible causes of increasing transnational terrorism.

Fisk (2019) tests the relationship between camp-settlement and a communal conflict in sub-Saharan Africa and proves that such conflict occurs more frequently in regions where refugees are camp-settled.

Finally, regarding the policies which destination countries should implement, Thielemann (2004) studies the evolution of asylum applications in Europe and suggests some policies so as to harmonize the refugee destination flows in European countries. Alfano et al. (2016) explore refugee policies in European destination countries and find that giving asylum to refugees does not cause negative consequences for the destination countries, or negative effects on the relationship with natives.

The second strand of literature focuses on women's economic and social rights, in both origin and destination countries. Bilgili et al. (2017) study the effects of displacement of

Congolese refugees in Rwanda in terms of food insecurity, subjective poverty and economic situation, focusing on gender. Ljungdell (1989), Baban et al. (2017) and Sansonetti (2016) analyze how to integrate female refugees in a safe way, while Powell et al. (2002) introduce female mutilation as a problem to deal with, and proposes policies for the integration of female refugees in United Kingdom. Torres (2018) remarks on the importance of feminist geopolitical perspectives on Latin American refugees. UNHCR (2003) analyzes sexual violence against refugees in both origin and destination countries. The UNHCR (2016) report on Age, Gender and Diversity Policy studies the causes and prevention of sexual and gender-based violence among women, LGTBI minorities and indigenous people. Martin-Storey et al. (2018) studies sexual violence in Canada across gender. Keygnaert et al. (2014) explore sexual gender-based violence in European countries against refugees, asylum seekers and undocumented migrants, and discusses determinants for 'Desirable Prevention'.

Bonewit and Shreeves (2016) describe the situation of female refugees and asylum-seekers in Germany from many perspectives, such as the services provided to them, the problems in reception centers, or the asylum procedure. Akram (2013) mentions some protection policies for female refugees in relation to the Millennium Development Goals.

As a contribution to this strand of literature, we analyze whether the respect for women's economic and social rights attract a greater proportion of female refugees in destination countries. On the other hand, we study whether the gender bias is greater (or lower) in destination countries isolated from origin countries where women do not have economic or social rights.

3.3 The data

In this chapter we use data on refugees by gender from the UNHCR, i.e. the percentages of male and female refugees, and the total number of refugees for each destination country and year for the period 2008-2016. We compile a longitudinal dataset following three steps. First, we multiply the number of total refugees in each destination country by the fraction of male and female refugees to obtain the total number of male and female refugees. Second, we compute the difference between the number of male and female refugees for each destination country and year. Third, we divide it by the total number of refugee to obtain the gender bias, our dependent variable. Next, we show some descriptive figures on the gender bias at the world-aggregate level. Figure 3.1 shows the difference between the number of male and female refugees per year in all destination countries. In general terms, the gender bias is always positive, ranging from a minimum of 200,000 more male refugees to a maximum of 1,200,000.

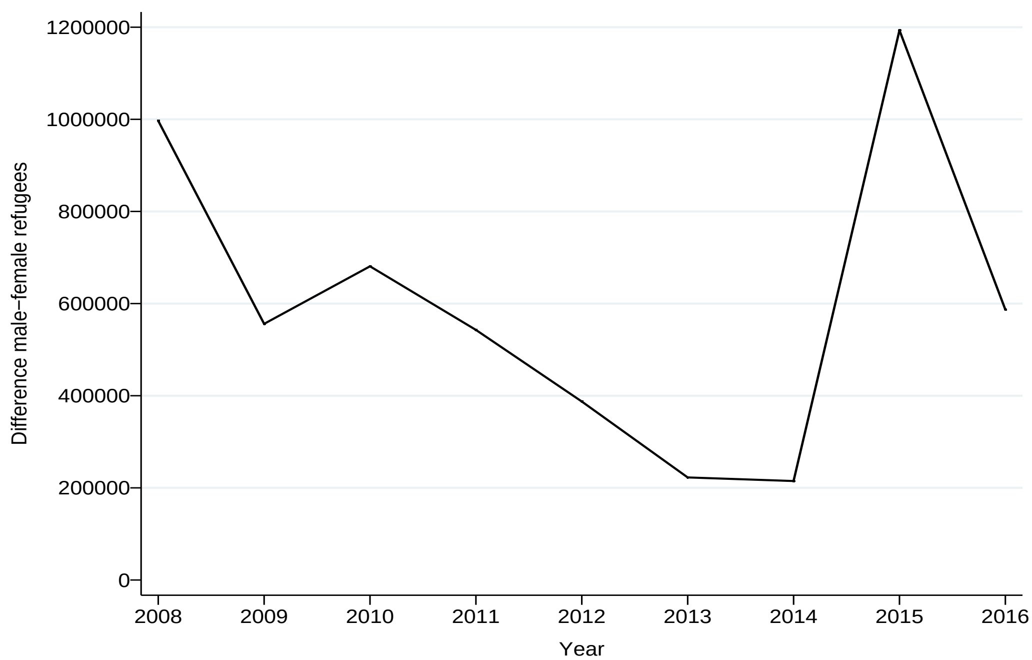

Figure 3.1: Gender bias in destination countries by year

Table 3.1 shows the average number of male and female refugees in the fifteen destination countries which host the largest refugee populations. It can be seen that the number of male refugees is greater than the number of female refugees in all countries but Jordan, Lebanon, Chad, Uganda and Sudan. In cases like France or Germany, the proportion of male refugees is 16 percent greater than the proportion of females.

Therefore, the existence of a gender bias in the number of refugees is clear. Figure 3.2

completes the picture by showing a histogram of the average gender bias of all destination countries for the period 2008-2016. The gender bias is negative in a very small proportion of destination countries. Most destination countries host more male refugees than females. In some countries, this difference is equal to one, indicating that all refugees are men.

Table 3.1: Average number of refugees: Top 15 destination countries (2008-2016)

	Males	Females	Difference	Percentage
Pakistan	879,443	768,388	111,055	6.74
Iran	546,643	417,459	129,184	13.39
Turkey	482,805	397,312	85,493	9.71
Jordan	262,631	270,829	-8,198	-1.54
Syria	267,487	258,838	8,649	1.64
Lebanon	236,483	246,429	-9,946	-2.1
Germany	279,606	200,646	78,960	16.44
Kenya	241,110	237,238	3,872	0.81
Ethiopia	204,438	200,691	3,747	0.92
Chad	166,220	212,143	-45,923	-12.14
Uganda	149,285	160,344	-11,059	-3,57
China	160,738	142,082	18,656	6,16
USA	143,408	124,957	18,451	6,85
France	135,234	97,559	37,675	16,47
Sudan	110,591	112,363	-1,772	-0,8

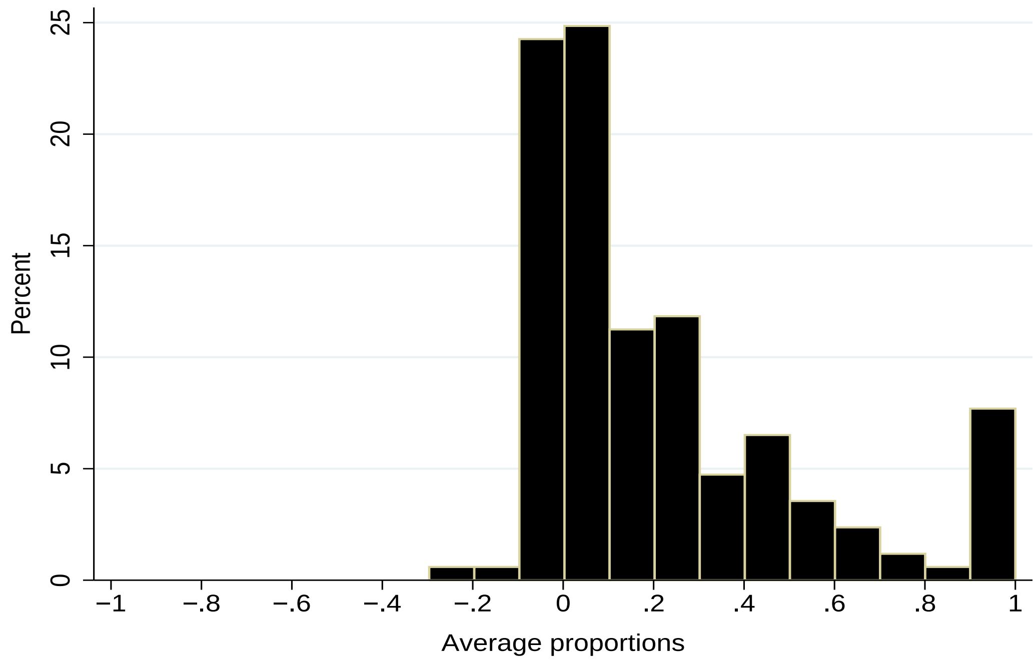

Figure 3.2: Histogram of average gender bias (2008-2016)

Our dataset includes destination-specific and country of origin-related covariates, along with a gender-related covariate. Regarding destination-specific covariates, the first covariate we add in our estimations is the civil liberties index, because forced migrants tend to go to countries where civil liberties are highly respected.

This index takes values between 1 (the highest level of civil liberties) and 7 (the lowest), respectively. Due to the lack of a sufficiently large number of observations at the seventh level, we merge levels 6 and 7. We want to determine whether more authoritarian countries host a greater proportion of male or female refugees.

A second covariate we include is the per capita GDP of destination countries. Previous studies have proven that refugees go to rich destination countries. Rich countries, in many cases, are also the countries with the highest civil liberty standards.

We want to determine whether the gender bias is positively or negatively related to the per capita GDP of destination countries. In addition, we also include the interaction between per capita GDP and the level of civil liberties in destination countries in our regressions.

The existence of armed conflicts in destination countries is the third covariate in our study of the gender bias in refugee flows. We add a categorical variable with three levels of conflict intensity: no conflict, low intensity conflicts and high intensity conflicts.

Many refugees move from their origin countries because of armed conflicts, and they tend to go to conflict-less countries. However, we do not know whether the relationship between armed conflicts in destination countries and the gender bias is positive or negative. This is what we want to determine with our estimations.

A fourth covariate accounted for in our analysis is the isolation of destination countries with respect to all origin countries. Isolation is a measure of how far a destination country is from all origin countries. Intuitively, a country geographically far (isolated) from origin countries would tend to have a lower number of refugees.

The isolation measure of each destination country is a function of the distance between a particular destination country and all other origin countries. We construct the isolation measure as follows. First, we normalize the value of the greatest distance between an origin and a destination country to one. Therefore, normalized distances are a proportion of the greatest, and range from zero to one.

We measure country j's isolation, I_j, as the sum of (normalized) distances from that particular destination country to all origin countries

$$I_j = \sum_{i=1}^{N} d_{ij}$$

where N is equal to the number of origin countries and d_{ij} represents the (normalized) distance from country i to country j. Intuitively, the isolation of a destination country with respect to origin countries is negatively related to refugee flows. However, we want to determine whether isolated destination countries host more male or female refugees.

Origin countries play a key role in determining refugee flows to destination countries. For example, it is reasonable to assume that poor and authoritarian origin countries generate more refugees than rich and free origin countries, and that refugees will tend to head for nearby destination countries. Therefore, we use distance as an aggregating factor of variables such as civil liberties and the per capita GDP of origin countries.

We compute weighted sums of countries of origin-specific variables weighted by their distances to a particular destination country, and repeat this for all destination countries. These variables are destination-specific variables, because they measure the distance from destination countries to rich/poor and free/authoritarian origin countries.

In particular, we compute a weighted sum of the levels of civil liberties in origin countries weighted by the distance to a particular destination country j as:

$$CL\_Origin_{jt} = \sum_{i=1}^{N} (d_{ij} * CL_{it})$$

where CL_i is the Civil Liberties level of origin country i, which is weighted by the distance between origin country i and destination country j. In order to understand this construct, consider a country very close to free origin countries.

This means a low value of this weighted sum (recall that the Civil Liberties index ranges from level 1, the highest civil liberty standards to 6, the lowest civil liberty standards). Similarly, a country very close to authoritarian origin countries would have a higher value of this weighted sum.

Intuition suggests that the country closer to authoritarian origin countries would tend to host a greater proportion of refugees, due to the proximity to countries from which refugees flee. Similarly, we compute a weighted sum of country of origin per capita GDP weighted by distance to each destination country,

$$GDP\_Origin_{jt} = \sum_{i=1}^{N} (d_{ij} * GDP_{it})$$

In this case, it seems reasonable to us to assume that a country close to poor origin countries, would host a greater proportion of refugees than a country close to rich origin countries.

However, it is not clear if the gender bias will be greater or lower in these destination countries.

In addition, we also compute a weighted sum of the interaction between per capita GDP and civil liberties, weighted by the distance to a destination country

$$GDP\_CL\_Origin_{jt} = \sum_{i=1}^{N}(d_{ij} * (\frac{GDP_{it}}{CL_{it}}))$$

In a similar way to the reasoning behind previous weighted sums, a small value of this magnitude corresponds to a destination country close to poor and authoritarian origin countries, while a large value would correspond to a destination country far from rich and free origin countries.

We point out that there are just a few cases of very rich and authoritarian, or very poor and free origin countries. Therefore, origin countries are generally rich and democratic, or poor and authoritarian. Intuition suggests that a country far from rich and free origin countries would host a greater proportion of refugees than a country far from poor and authoritarian origin countries.

While the inclusion of civil liberties in our analysis seems reasonable, this variable measures the rights of both men and women altogether, without taking into account whether male and female civil liberties are equally respected.

To account for this shortcoming of the civil liberties measure, we also consider a gender-related variable in our analysis: the World Bank Women, Business and Law index ($WBLI$). This index is computed as an average of the following factors: women's rights to travel to places, start a job, get paid, get married, have children, run a business, manage assets and receive a pension.

The $WBLI$ takes values from zero (the absence of economic, social and judicial rights) to 100 (the highest level). Arguably, holding everything else constant, women should go to destination countries where their rights are more respected.

In many origin countries, women do not have the same economic or social rights as men.

Therefore, in addition to women's rights in destination countries, women's rights in origin countries are also very important in the determination of refugee destination flows.

As we explained above, we compute a weighted sum of the $WBLI$ in origin countries weighted by their distances to a destination country as follows:

$$WBLI\_Origin_j = \sum_{i=1}^{N} (d_{ij} * WBLI_i)$$

High values of this index indicate isolation from origin countries where women's rights are respected. Intuition suggests that the gender bias should be larger in destination countries far from origin countries where women's rights are respected, because women would decide not to go to such destination countries.

Finally, we have also computed a weighted sum of the interaction of per capita GDP and the WBLI at origin countries weighted by the distance to a destination country:

$$GDP\_WBLI\_Origin_j = \sum_{i=1}^{N} (d_{ij} * GDP_i * WBLI_i)$$

Intuition suggests that destination countries far from origin countries with high levels of per capita GDP and rights for women (or closer to poor and authoritarian origin countries, from where refugees flee), attract a greater proportion of male refugees.

We also introduce a variable which measures the total number of origin countries which suffer armed conflicts and are contiguous to each destination country. Intuition suggests that a destination country contiguous to a considerable number of origin countries in conflict will host a greater proportion of refugees. However, we cannot determine the effect on the gender bias without including this variable in our estimations.

Variable definitions and sources are shown in Appendix 3.6, and summary statistics are reported in Table 3.2 without logarithms and in Table 3.3 with logarithms. The gender bias ranges from -0.54 to one. This means that, in a specific destination country, the bias is 54% greater for female refugees and in another destination country, there are only male refugees. In 108 observations of our dataset the bias is equal to zero. Destination countries with the lowest respect for civil liberties are the most frequent in the sample, while countries with the greatest respect for civil liberties are the least frequent observations in the sample. The isolation of destination countries has a mean value equal to 90.17, which is closer to its minimum value than to its maximum value. The GDP per capita of destination countries ranges from its minimum value, which equals 568.08 dollars, to its maximum value, which equals 129,349.9 dollars, with an average value of 15,237.9 dollars. There are no armed conflicts in 81 percent of observations of this sample and, amongst armed conflicts, low intensity conflicts have a greater frequency than high intensity conflicts. Most destination countries have a moderately high WBLI score, as can be seen in its mean value.

Table 3.2: Summary statistics of female and male refugees and covariates (without logarithms)

Variable	# obs.	mean	stand.dev.	min	max
Gender bias	1205	0.17	0.26	-0.54	1
Civil Liberties level 1	1166	0.12	0.33	0	1
Civil Liberties level 2	1166	0.14	0.35	0	1
Civil Liberties level 3	1166	0.18	0.38	0	1
Civil Liberties level 4	1166	0.17	0.38	0	1
Civil Liberties level 5	1166	0.19	0.39	0	1
Civil Liberties levels 6 & 7	1166	0.19	0.39	0	1
Isolation	1189	90.17	15.90	72.90	148.92
GDPpcDestination	1132	15,237.90	19,248.94	568.08	129,349.90
Conflict Level 0	1205	0.81	0.39	0	1
Conflict Level 1	1205	0.14	0.35	0	1
Conflict Level 2	1205	0.05	0.21	0	1
WBLI	1015	69.39	18.15	23.13	100
Contiguous Origin Countries Conflict	1205	0.94	1.19	0	6

Table 3.3: Summary statistics of female and male refugees and covariates (with logarithms)

Variable	# obs.	mean	stand.dev.	min	max
Gender bias	1205	0.17	0.26	-0.54	1
Civil Liberties level 1	1166	0.12	0.33	0	1
Civil Liberties level 2	1166	0.14	0.35	0	1
Civil Liberties level 3	1166	0.18	0.38	0	1
Civil Liberties level 4	1166	0.17	0.38	0	1
Civil Liberties level 5	1166	0.19	0.39	0	1
Civil Liberties levels 6 & 7	1166	0.19	0.39	0	1
(Log) Isolation	1189	4.49	0.17	4.29	5.01
(Log) GDPpcDestination	1132	8.96	1.22	6.34	11.77
Conflict Level 0	1205	0.81	0.39	0	1
Conflict Level 1	1205	0.14	0.35	0	1
Conflict Level 2	1205	0.05	0.21	0	1
(Log) WBLI	1015	4.20	0.32	3.14	4.61
Contiguous Origin Countries Conflict	1205	0.94	1.19	0	6

3.4 Empirics

3.4.1 Gender bias without gender-related covariates

Table 3.4 reports marginal effects of OLS linear regression estimates with standard errors clustered at country level using a total of 1,095 observations from 154 countries. Regression estimations are included in Table 3.8 in the Appendix.

As we introduce an interaction term, the effect of civil liberties and per capita GDP can only be determined by computing their marginal effects. These marginal effects compare each level of the civil liberties index with respect to the reference, which is the greatest respect of civil liberties.

Regarding the per capita GDP of destination countries, we compute its marginal effect at each civil liberties level, from one to six. Column 1 estimates suggest that the gender bias diminishes as civil liberties deteriorate. This result is counter-intuitive, as women would enjoy more freedom in countries with high standards of civil liberties.

In addition, the gender bias increases in richer and more isolated countries. Armed conflict is not a significant determinant of the gender bias. Column 2 includes the total number of contiguous origin countries which suffer armed conflicts, which turns out to be significant, and has a negative coefficient.

This means that the gender bias is lower in destination countries with more contiguous origin countries in conflict. Summing up, it can be observed that a lower level of respect for civil liberties implies a lower gender bias, so results related to civil liberties are counter-intuitive, as in Column 1.

In addition, the per capita GDP affects the gender bias positively, as shown by the marginal effects of per capita GDP in Table 3.4 for each level of the civil liberties index. Regarding isolation, its interpretation has not changed: a destination country which is more isolated with respect to the rest of origin countries attracts a greater proportion of male refugees.

The gender bias is not different in countries with armed conflicts (both intensity levels) with respect to countries where armed conflicts do not occur.Columns 3, 4 and 5 add the marginal effects corresponding to the country of origin-related variables weighted by distance.

With respect to the results reported in Column 2, adding the origin-related variables does not quantitatively change the results of the other covariates to any significant effect. Civil liberties consistently affect the gender bias negatively, while the per capita GDP of the destination country maintains a positive effect on the gender bias, as marginal effects show in Table 3.4. The results for civil liberties are remarkably robust, despite the different specifications across columns. The total number of contiguous origin countries in conflict for each destination country consistently has a negative effect on the gender bias. Amongst the origin-related variables, the per capita GDP of origin countries weighted by distance is significant in Columns 3 and 4, but not in Column 5. Destination countries isolated from rich origin countries (or close to poor origin countries) receive a greater proportion of female refugees. In Column 5, the interaction of per capita GDP and civil liberties of origin countries weighted by distance is significant and negative. This means that destination countries far from rich and free origin countries host a greater proportion of female refugees than destination countries close to free origin countries.

Table 3.4: Marginal effects of OLS regressions for the gender bias (2008-2016)

	(1)	(2)	(3)	(4)	(5)
Civil Liberties level 2	-0.3166*	-0.3198*	-0.3156**	-0.3101**	-0.3241**
	(0.1696)	(0.1705)	(0.1543)	(0.1556)	(0.1531)
Civil Liberties level 3	-0.3563**	-0.3578**	-0.3717**	-0.3749**	-0.3858**
	(0.1687)	(0.1694)	(0.1525)	(0.1544)	(0.1525)
Civil Liberties level 4	-0.3541**	-0.3512**	-0.3733**	-0.3640**	-0.3678**
	(0.1688)	(0.1697)	(0.1527)	(0.1543)	(0.1517)
Civil Liberties level 5	-0.4414***	-0.4373***	-0.4475***	-0.4325***	-0.4159***
	(0.1686)	(0.1695)	(0.1532)	(0.1548)	(0.1529)
Civil Liberties level 6	-0.4393***	-0.4223**	-0.4424***	-0.4273***	-0.4102***
	(0.1686)	(0.1701)	(0.1537)	(0.1549)	(0.1529)
(Log) Isolation	0.2157**	0.1905*	0.9747***	0.7864*	3.2512***
	(0.1058)	(0.1059)	(0.3346)	(0.4140)	(1.1487)
Conflict Intensity 1	0.0192	0.0302	0.0264	0.0273	0.0271
	(0.0351)	(0.0344)	(0.0336)	(0.0325)	(0.0308)
Conflict Intensity 2	0.0319	0.0375	0.0140	0.0155	0.0277
	(0.0310)	(0.0278)	(0.0318)	(0.0327)	(0.0302)
(Log) GDPpc ǀ CL1	-0.1602*	-0.1677*	-0.1935**	-0.2034**	-0.2379***
	(0.0959)	(0.0963)	(0.0869)	(0.0895)	(0.0888)
(Log) GDPpc ǀ CL2	0.0999***	0.0997***	0.0791**	0.0709*	0.0737*
	(0.0372)	(0.0387)	(0.0375)	(0.0386)	(0.0392)
(Log) GDPpc ǀ CL3	0.0604**	0.060**	0.0424*	0.0286	0.0270
	(0.0254)	(0.0258)	(0.0275)	(0.0299)	(0.0307)
(Log) GDPpc ǀ CL4	0.0796***	0.0786***	0.0534***	0.0493**	0.0467**
	(0.0168)	(0.0170)	(0.0204)	(0.0206)	(0.0191)
(Log) GDPpc ǀ CL5	0.0066	0.0044	-0.0119	-0.0128	-0.0052
	(0.0179)	(0.0185)	(0.0192)	(0.0186)	(0.0176)
(Log) GDPpc ǀ CL6	0.0329***	0.0317**	0.0101	0.0060	0.0139
	(0.0113)	(0.0126)	(0.0141)	(0.0140)	(0.0129)
Contiguous Countries Conflict		-0.0180**	-0.0215**	-0.0204**	-0.0177**
		(0.0086)	(0.0088)	(0.0088)	(0.0085)
(Log) GDPpc Origin			-0.5134**	-0.5970***	0.5449
			(0.2075)	(0.2096)	(0.4180)
(Log) CL Origin				0.2273	-1.1453
				(0.2225)	(0.5949)
(Log) CL GDPpc Origin					-1.3137**
					(0.5129)
Constant	0.9729	1.1692	5.0205***	5.8775***	4.1247**
	(1.2022)	(1.2064)	(1.8524)	(1.9813)	(1.8640)
Observations	1,095	1,095	1,095	1,095	1,095
Number of countries	154	154	154	154	154
R-squared	0.2169	0.2225	0.2463	0.2491	0.2684

Standard errors have been computed with the Delta Method. *** $p<0.01$, ** $p<0.05$, * $p<0.1$
All regressions include year dummies.

3.4.2 Gender bias with gender-related variables: the WBL index

In this subsection, we study the effect of women's specific rights on the gender bias. Table 3.5 reports marginal effects of OLS linear regression estimates where the specification includes isolation, the conflict intensity indicator and the per capita GDP.

Then, the specification replaces civil liberties with the WBLI; it also includes the interaction between the WBLI and the per capita GDP of destination countries. Regression estimations are included in Table 3.9 in the Appendix.

In order to compute the marginal effects of the interacted variables, we divide them in percentiles and we show the results of the 20, 40, 60 and 80^{th} percentiles. As explained above, the civil liberties index does not take into account whether there is a gender difference in civil liberties.

To account for this possibility, in this section we introduce the WBLI. As a result, the sample is now reduced to the period 2009-2016 period and 151 countries for a total of 968 observations. Column 1 shows that the gender bias is positively related to higher levels of women's economic and social rights, except for the poorest destination countries, where the WBLI has no effect on the gender bias.

Again, this result is counter-intuitive.If female refugees had the opportunity, they would choose to go to countries where their economic rights were respected. Armed conflict in destination countries is not relevant when explaining the gender bias.

More isolated destination countries host a greater proportion of male refugees. Finally, the gender bias is greater in rich destination countries at every level of the WBLI. Column 2 includes the total number of contiguous origin countries, where there are armed conflicts, of each destination country.

It turns out that the gender bias is lower in destination countries contiguous to origin countries where there are armed conflicts, while the effect of the other covariates remain qualitatively similar to those in Column 1.

Table 3.5: Marginal Effects with gender-related variables for the gender bias (2009-2016)

	(1)	(2)	(3)	(4)	(5)
(Log) WBLI | P_{20}GDPpc	0.0539	0.0245	0.0245	0.0240	0.0316
	0.0488)	(0.0539)	(0.0537)	(0.0505)	(0.0506)
(Log) WBLI | P_{40} GDPpc	0.1012***	0.0770*	0.0755*	0.0581	0.0427
	(0.0381)	(0.0425)	(0.0424)	(0.0398)	(0.0419)
(Log) WBLI | P_{60} GDPpc	0.1330***	0.1123***	0.1097***	0.0809**	0.0502
	(0.0372)	(0.0412)	(0.0411)	(0.0398)	(0.0457)
(Log) WBLI | P_{80} GDPpc	0.1635***	0.1462***	0.1425***	0.1028**	0.0573
	(0.0415)	(0.0455)	(0.0453)	(0.0454)	(0.0554)
(Log) Isolation	0.3366***	0.3081***	0.9006***	5.2483**	4.7845**
	(0.1021)	(0.1018)	(0.3416)	(2.2601)	(2.3741)
Conflict Level 1	-0.0094	0.0026	-0.0023	0.0029	0.0062
	(0.0346)	(0.0341)	(0.0342)	(0.0327)	(0.0309)
Conflict Level 2	0.0047	0.0114	-0.0087	0.0108	0.0152
	(0.0341)	(0.0951)	(0.0950)	(0.0379)	(0.0388)
(Log) GDPpc | P_{20} WBLI	0.0496***	0.0457***	0.0313**	0.0369***	0.0303**
	(0.009)	(0.01)	(0.0131)	(0.0122)	(0.0145)
(Log) GDPpc | P_{40} WBLI	0.0616***	0.0591***	0.0442***	0.0455***	0.0331*
	(0.01)	(0.0105)	(0.0134)	(0.0132)	(0.0181)
(Log) GDPpc | P_{60} WBLI	0.0652***	0.0631***	0.0481***	0.0481***	0.0340*
	(0.0109)	(0.0113)	(0.0140)	(0.0140)	(0.0195)
(Log) GDPpc | P_{80} WBLI	0.0695***	0.0679***	0.0527***	0.0512***	0.0350
	(0.0122)	(0.0125)	(0.0150)	(0.0152)	(0.0215)
Contiguous Countries Conflict		-0.0185*	-0.0230**	-0.0179*	-0.0146
		(0.0101)	(0.0107)	(0.0104)	(0.01)
(Log) GDP Origin			-0.3923*	-0.3306*	0.0884
			(0.2116)	(0.1902)	(0.3974)
(Log) WBLI Origin				-3.4573**	-2.7799
				(1.6616)	(1.8773)
(Log) WBLI GDP Origin					-0.6309
					(0.5044)
Constant	-0.5706	-0.1160	2.7422	11.0304**	12.3199**
	(0.9609)	(1.0350)	(1.8278)	(5.0207)	(4.8363)
Observations	968	968	968	968	968
Number of countries	151	151	151	151	151
R-squared	0.1826	0.1883	0.2024	0.2197	0.2253

Standard errors have been computed with the Delta Method
*** p<0.01, ** p<0.05, * p<0.1. All regressions include year dummies.

Column 3 includes the per capita GDP of origin countries weighted by distance. It turns out that the gender bias is lower in destination countries isolated from rich origin countries (or close to poor countries) while the effect of the other covariates remains qualitatively similar to that in Column 2.

Column 4 includes the WBLI of origin countries weighted by distance. The gender bias is lower in destination countries isolated from origin countries with a high WBLI (close to countries with a low WBLI). As in Column 3, more isolated destination countries tend to exhibit a greater gender bias, but now the marginal effect is much larger. As in Column 3, GDP at origin is significant.Column 5 adds the interaction between the per capita GDP and the WBLI of origin countries weighted by distance. This interaction turns out to be insignificant. Meanwhile, the WBLI of origin countries and the interaction between per capita GDP and WBLI of origin countries weighted by distance are no longer significant. Therefore, our preferred estimation is Column 4. Summarizing the results of Tables 3.4 and 3.5, the gender bias tends to be larger in isolated and rich countries with high standards of civil liberties and women's rights. A feasible reading of this result is that female refugees are discriminated against and therefore are forced to flee to nearby poor countries with low civil liberty standards countries.

3.4.3 Robustness checks

To account for the presence of unobserved confounders Tables 3.6 and 3.7 report marginal effects of random effects estimates without and with gender-related variables respectively. Regression estimations are included in Tables 3.10 and 3.11 in the Appendix. The results obtained are qualitatively the same as those reported above. Random effects estimates might be biased if the unobserved confounders are correlated with the covariates included in the regression. Fixed effects estimates would not suffer from such a bias.However, as some of the key explanatory variables, namely civil liberties and women's rights, vary very little with time, their effect on the gender bias cannot be identified.

Table 3.6: Marginal effects of random effects regressions for the gender bias (2008-2016)

	(1)	(2)	(3)	(4)	(5)
Civil Liberties level 2	-0.3552**	-0.3544**	-0.3567**	-0.3531**	-0.3434**
	(0.1469)	(0.1461)	(0.1412)	(0.1414)	(0.1426)
Civil Liberties level 3	-0.4046***	-0.4010***	-0.4098***	-0.4074***	-0.3819***
	(0.1482)	(0.1473)	(0.1423)	(0.1427)	(0.1447)
Civil Liberties level 4	-0.4291***	-0.4205***	-0.4331***	-0.4247***	-0.4001***
	(0.1502)	(0.1494)	(0.1437)	(0.1439)	(0.1471)
Civil Liberties level 5	-0.4804***	-0.4686***	-0.4816***	-0.4691***	-0.4416***
	(0.1502)	(0.1495)	(0.1443)	(0.1449)	(0.1477)
Civil Liberties level 6	-0.4795***	-0.4645***	-0.4820***	-0.4674***	-0.4345***
	(0.1523)	(0.1521)	(0.1473)	(0.1474)	(0.1503)
(Log) Isolation	0.2749**	0.2444**	0.9033**	0.6846	2.3383**
	(0.1186)	(0.1176)	(0.4004)	(0.5249)	(1.1291)
Conflict Level 1	-0.0033	-0.0069	-0.0080	-0.0070	-0.0073
	(0.0173)	(0.0178)	(0.0178)	(0.0176)	(0.0174)
Conflict Level 2	-0.0163	-0.0167	-0.0212	-0.0200	-0.0221
	(0.0453)	(0.0458)	(0.0473)	(0.0473)	(0.0470)
(Log) GDPpc | CL1	-0.2058**	-0.2109**	-0.2301***	-0.2399***	-0.2601***
	(0.0846)	(0.0840)	(0.0814)	(0.0841)	(0.0838)
(Log) GDPpc | CL2	0.0620*	0.0601*	0.0478	0.0388	0.0318
	(0.0319)	(0.0329)	(0.0325)	(0.0343)	(0.0344)
(Log) GDPpc | CL3	0.0580**	0.0561**	0.0438*	0.0347	0.0321
	(0.0234)	(0.0232)	(0.0258)	(0.0273)	(0.0281)
(Log) GDPpc | CL4	0.0492**	0.0488**	0.0330	0.0279	0.0261
	(0.02)	(0.02)	(0.0222)	(0.0225)	(0.0219)
(Log) GDPpc | CL5	0.0285*	0.0252	0.0096	0.0067	0.0079
	(0.0165)	(0.0171)	(0.0183)	(0.0187)	(0.0179)
(Log) GDPpc | CL6	0.0161	0.0145	-0.0022	-0.0055	-0.0038
	(0.0158)	(0.0159)	(0.0181)	(0.0187)	(0.0184)
Contiguous Countries Conflict		-0.0186**	-0.0197**	-0.0186**	-0.0183**
		(0.0078)	(0.0078)	(0.0076)	(0.0076)
(Log) GDPpc Origin			-0.4421*	-0.5259**	0.2451
			(0.2531)	(0.2460)	(0.4300)
(Log) CL Origin				0.2501	-0.6094
				(0.2899)	(0.5909)
(Log) CL GDP Origin					-0.9464**
					(0.4773)
Constant	1.1966	1.3936	4.7503**	5.6243***	4.8734**
	(1.1439)	(1.1365)	(2.1025)	(2.1760)	(2.1884)
Observations	1,095	1,095	1,095	1,095	1,095
Number of countries	154	154	154	154	154
R-squared	0.1990	0.2030	0.2314	0.2340	0.2501

Standard errors have been computed with the Delta Method. *** p<0.01, ** p<0.05, * p<0.1
All regressions include year dummies.

Table 3.7: Marginal Effects of random-effects regressions, with gender-related variables, for the gender bias (2009-2016)

	(1)	(2)	(3)	(4)	(5)
(Log) WBLI ∣ P_{20}GDPpc	0.0539	0.0245	0.0245	0.0240	0.0316
	0.0488)	(0.0539)	(0.0537)	(0.0505)	(0.0506)
(Log) WBLI ∣ P_{40} GDPpc	0.1012***	0.0770*	0.0755*	0.0581	0.0427
	(0.0381)	(0.0425)	(0.0424)	(0.0398)	(0.0419)
(Log) WBLI ∣ P_{60} GDPpc	0.1330***	0.1123***	0.1097***	0.0809**	0.0502
	(0.0372)	(0.0412)	(0.0411)	(0.0398)	(0.0457)
(Log) WBLI ∣ P_{80} GDPpc	0.1635***	0.1462***	0.1425***	0.1028**	0.0573
	(0.0415)	(0.0455)	(0.0453)	(0.0454)	(0.0554)
(Log) Isolation	0.3366***	0.3081***	0.9006***	5.2483**	4.7845**
	(0.1021)	(0.1018)	(0.3416)	(2.2601)	(2.3741)
Conflict Level 1	-0.0094	0.0026	-0.0023	0.0029	0.0062
	(0.0346)	(0.0341)	(0.0342)	(0.0327)	(0.0309)
Conflict Level 2	0.0047	0.0114	-0.0087	0.0108	0.0152
	(0.0341)	(0.0951)	(0.0950)	(0.0379)	(0.0388)
(Log) GDPpc ∣ P_{20} WBLI	0.0496***	0.0457***	0.0313**	0.0369***	0.0303**
	(0.009)	(0.01)	(0.0131)	(0.0122)	(0.0145)
(Log) GDPpc ∣ P_{40} WBLI	0.0616***	0.0591***	0.0442***	0.0455***	0.0331*
	(0.01)	(0.0105)	(0.0134)	(0.0132)	(0.0181)
(Log) GDPpc ∣ P_{60} WBLI	0.0652***	0.0631***	0.0481***	0.0481***	0.0340*
	(0.0109)	(0.0113)	(0.0140)	(0.0140)	(0.0195)
(Log) GDPpc ∣ P_{80} WBLI	0.0695***	0.0679***	0.0527***	0.0512***	0.0350
	(0.0122)	(0.0125)	(0.0150)	(0.0152)	(0.0215)
Contiguous Countries Conflict		-0.0185*	-0.0230**	-0.0179*	-0.0146
		(0.0101)	(0.0107)	(0.0104)	(0.01)
(Log) GDP Origin			-0.3923*	-0.3306*	0.0884
			(0.2116)	(0.1902)	(0.3974)
(Log) WBLI Origin				-3.4573**	-2.7799
				(1.6616)	(1.8773)
(Log) WBLI GDP Origin					-0.6309
					(0.5044)
Constant	-0.5706	-0.1160	2.7422	11.0304**	12.3199**
	(0.9609)	(1.0350)	(1.8278)	(5.0207)	(4.8363)
Observations	968	968	968	968	968
Number of countries	151	151	151	151	151
R-squared	0.1826	0.1883	0.2024	0.2197	0.2253

Standard errors have been computed with the Delta Method. *** p<0.01, ** p<0.05, * p<0.1.
All regressions include year dummies.

3.5 Conclusions

This chapter studies the gender bias (the difference in the number of male and female refugees) in destination flows for the period 2008-2016. This chapter takes into account several determinants of refugee destination flows considered in previous studies.

The contribution of this chapter is to analyze the effect of these factors on the gender bias, which has never been dealt with before. We analyze the effect of various types of covariates on the gender bias: destination-specific variables, country of origin-related variables and gender-related variables.

Summing up the results obtained, we find that the gender bias in refugee destinations is larger in countries with high levels of economic and social rights, in isolated destination countries and also in rich destination countries.As regards origin-related variables, the bias is larger in destination countries isolated from poor and authoritarian origin countries.

An interpretation of this result is that men might have the necessary economic and social rights in their origin countries to flee to free destination countries. Women, on the contrary, lacking economic and social rights can only escape to closer destination countries which often happen to be authoritarian and poor.

3.6 Appendix D: Data sources of Chapter 3

The data set compiled is a blend of the following data sets:

1. Refugees by gender Dataset. United Nations High Commissioner for the Refugees (UNHCR). Time frame: 2008-2016. Variables used: Gender bias
   Availability: `http://www.unhcr.org/statistical-yearbooks.html`

2. Armed Conflict Dataset. Uppsala Conflict Data Program (UCDP) Uppsala University / Peace Research Institute Oslo (PRIO). Time frame: 1946-2017. Variables used: Location. The name(s) of the country/countries whose government(s) have a primary claim to the issue in dispute. Year of observation. The date when the conflict activity reached 25 battle-related deaths in a year. The date when conflict activity ended. The intensity level of the armed conflict, total number of contiguous origin countries in conflict.
   Availability: `https://www.prio.org/Data/Armed-Conflict/UCDP-PRIO/`

3. Freedom in the World Dataset. Freedom House. Time frame: 1972-2017. Variables used: Civil Liberties. Availability: `https://freedomhouse.org`

4. Geographical and Distance dataset. CEPII, SciencesPo Department of Economics. Dyadic data set. Variables used: Simple distance (most populated cities, km). Transformed into longitudinal variables: isolation, the sum of total distance w.r.t. the rest of origin countries, country of origin-related variables weighted by distance, total number of contiguous origin countries in conflict.
   Availability: `http://econ.sciences-po.fr/staff/thierry-mayer`

5. World Development Indicators. The World Bank.Variables used: GDP per capita, PPP (current international dollars). Availability: `http://databank.worldbank.org/data/home.aspx`

6. World Bank. Women, Business and the Law Dataset. Variables used: Women, Business and the Law index (WBLINDEX). Availability: `http://wbl.worldbank.org/`

3.7 Appendix E: Parameter estimates of Chapter 3

Table 3.8: Parameter estimates corresponding to Table 3.4

	(1)	(2)	(3)	(4)	(5)
Civil Liberties level 2	-2.6416**	-2.7102**	-2.7530***	-2.7618***	-3.1090***
	(1.0947)	(1.1040)	(0.9973)	(1.0118)	(1.0037)
Civil Liberties level 3	-2.3279**	-2.3928**	-2.4808***	-2.4487**	-2.7538***
	(1.0406)	(1.0467)	(0.9505)	(0.9606)	(0.9493)
Civil Liberties level 4	-2.4975**	-2.5527**	-2.5802***	-2.6230***	-2.9118***
	(1.0365)	(1.0410)	(0.9431)	(0.9610)	(0.9437)
Civil Liberties level 5	-1.9326*	-1.9757*	-2.0713**	-2.1360**	-2.4960***
	(1.0343)	(1.0387)	(0.9365)	(0.9524)	(0.9427)
Civil Liberties level 6	-2.1648**	-2.2045**	-2.2624**	-2.2988**	-2.6610***
	(1.0239)	(1.0294)	(0.9271)	(0.9429)	(0.9308)
(Log) Isolation	0.2157**	0.1905*	0.9747***	0.7864*	3.2512***
	(0.1058)	(0.1059)	(0.3346)	(0.4140)	(1.1487)
(Log) GDPpc Destination	-0.1602*	-0.1677*	-0.1935**	-0.2034**	-0.2379***
	(0.0959)	(0.0963)	(0.0869)	(0.0895)	(0.0888)
Conflict Intensity 1	0.0192	0.0302	0.0264	0.0273	0.0271
	(0.0351)	(0.0344)	(0.0336)	(0.0325)	(0.0308)
Conflict Intensity 2	0.0319	0.0375	0.0140	0.0155	0.0277
	(0.0310)	(0.0278)	(0.0318)	(0.0327)	(0.0302)
(Log) GDPpc*CL level 2	0.2601**	0.2674**	0.2727***	0.2743***	0.3115***
	(0.1053)	(0.1063)	(0.0964)	(0.0980)	(0.0978)
(Log) GDPpc*CL level 3	0.2206**	0.2277**	0.2359**	0.2320**	0.2649***
	(0.0991)	(0.0998)	(0.0910)	(0.0919)	(0.0909)
(Log) GDPpc*CL level 4	0.2398**	0.2463**	0.2469***	0.2527***	0.2846***
	(0.0981)	(0.0986)	(0.0896)	(0.0917)	(0.0901)
(Log) GDPpc*CL level 5	0.1668*	0.1721*	0.1816**	0.1906**	0.2327**
	(0.0975)	(0.0979)	(0.0884)	(0.0902)	(0.0896)
(Log) GDPpc*CL level 6	0.1930**	0.1994**	0.2036**	0.2094**	0.2518***
	(0.0963)	(0.0969)	(0.0873)	(0.0891)	(0.0882)
Contiguous Countries Conflict		-0.0180**	-0.0215**	-0.0204**	-0.0177**
		(0.0086)	(0.0088)	(0.0088)	(0.0085)
(Log) GDP Origin			-0.5134**	-0.5970***	0.5449
			(0.2075)	(0.2096)	(0.4180)
(Log) CL Origin				0.2273	-1.1453
				(0.2225)	(0.5949)
(Log) CL GDP Origin					-1.3137**
					(0.5129)
Constant	0.9729	1.1692	5.0205***	5.8775***	4.1247**
	(1.2022)	(1.2064)	(1.8524)	(1.9813)	(1.8640)
Observations	1,095	1,095	1,095	1,095	1,095
Number of countries	154	154	154	154	154
R-squared	0.2169	0.2225	0.2463	0.2491	0.2684

Country-clustered robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
All regressions include year dummies.

Table 3.9: Parameter estimates corresponding to Table 3.5

	(1)	(2)	(3)	(4)	(5)
(Log) WBLI	-0.3012	-0.3697*	-0.3578*	-0.2311	-0.0517
	(0.2007)	(0.2168)	(0.2156)	(0.2145)	(0.2412)
(Log) Isolation	0.3366***	0.3081***	0.9006***	5.2483**	4.7845**
	(0.1021)	(0.1018)	(0.3416)	(2.2601)	(2.3741)
Conflict Level 1	-0.0094	0.0026	-0.0023	0.0029	0.0062
	(0.0346)	(0.0341)	(0.0342)	(0.0327)	(0.0309)
Conflict Level 2	0.0047	0.0114	-0.0087	0.0108	0.0152
	(0.0341)	(0.0344)	(0.0414)	(0.0379)	(0.0388)
(Log) GDPpc Destination	-0.1349	-0.1591*	-0.1674*	-0.0957	-0.0130
	(0.0873)	(0.0951)	(0.0950)	(0.0946)	(0.1048)
(Log) WBLI * (Log) GDPpc	0.0462**	0.0513**	0.0497**	0.0332	0.0108
	(0.0215)	(0.0231)	(0.0230)	(0.0233)	(0.0272)
Contiguous Countries Conflict		-0.0184*	-0.0226**	-0.0179*	-0.0146
		(0.0101)	(0.0107)	(0.0104)	(0.0100)
(Log) GDP Origin			-0.3923*	-0.3306*	0.0884
			(0.2116)	(0.1902)	(0.3974)
(Log) WBLI Origin				-3.4573**	-2.7799
				(1.6616)	(1.8774)
(Log) WBLI GDP Origin					-0.6309
					(0.5044)
Constant	-0.5705	-0.1160	2.7422	11.0304**	12.3199**
	(0.9609)	(1.0350)	(1.8278)	(5.0207)	(4.8363)
Observations	968	968	968	968	968
Number of countries	151	151	151	151	151
R-squared	0.1826	0.1883	0.2024	0.2197	0.2253

Country-clustered robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
All regressions include year dummies.

Table 3.10: Parameter estimates corresponding to Table 3.6

	(1)	(2)	(3)	(4)	(5)
Civil Liberties level 2	-2.7484***	-2.7774***	-2.8405***	-2.8446***	-2.9600***
	(0.9829)	(0.9810)	(0.9454)	(0.9509)	(0.9541)
Civil Liberties level 3	-2.7623***	-2.7881***	-2.8574***	-2.8618***	-3.0080***
	(0.9283)	(0.9226)	(0.8918)	(0.8962)	(0.8955)
Civil Liberties level 4	-2.7083***	-2.7418***	-2.7842***	-2.8187***	-2.9657***
	(0.9288)	(0.9219)	(0.8939)	(0.9041)	(0.9012)
Civil Liberties level 5	-2.5742***	-2.5798***	-2.6236***	-2.6732***	-2.8371***
	(0.9124)	(0.9062)	(0.8810)	(0.8896)	(0.8883)
Civil Liberties level 6	-2.4627***	-2.4792***	-2.5187***	-2.5626***	-2.7306***
	(0.9085)	(0.9027)	(0.8763)	(0.8854)	(0.8838)
(Log) Isolation	0.2749**	0.2444**	0.9033**	0.6846	2.3383**
	(0.1186)	(0.1176)	(0.4004)	(0.5249)	(1.1291)
(Log) GDPpc Destination	-0.2058**	-0.2109**	-0.2300***	-0.2399***	-0.2600***
	(0.0846)	(0.0840)	(0.0814)	(0.0841)	(0.0838)
Conflict Level 1	-0.0033	-0.0069	-0.0080	-0.0070	-0.0073
	(0.0173)	(0.0178)	(0.0178)	(0.0176)	(0.0174)
Conflict Level 2	-0.0163	-0.0167	-0.0212	-0.0200	-0.0221
	(0.0453)	(0.0458)	(0.0473)	(0.0473)	(0.0470)
(Log) GDPpc*CL level 2	0.2677***	0.2711***	0.2779***	0.2787***	0.2918***
	(0.0946)	(0.0945)	(0.0911)	(0.0918)	(0.0922)
(Log) GDPpc*CL level 3	0.2637***	0.2670***	0.2738***	0.2746***	0.2921***
	(0.0887)	(0.0882)	(0.0853)	(0.0858)	(0.0856)
(Log) GDPpc*CL level 4	0.2550***	0.2597***	0.2630***	0.2678***	0.2860***
	(0.0883)	(0.0876)	(0.0851)	(0.0865)	(0.0860)
(Log) GDPpc*CL level 5	0.2342***	0.2362***	0.2396***	0.2466***	0.2679***
	(0.0861)	(0.0855)	(0.0832)	(0.0844)	(0.0841)
(Log) GDPpc*CL level 6	0.2218***	0.2254***	0.2278***	0.2344***	0.2562***
	(0.0856)	(0.0851)	(0.0827)	(0.0840)	(0.0837)
Contiguous Countries Conflict		-0.0186**	-0.0197**	-0.0186**	-0.0183**
		(0.0078)	(0.0077)	(0.0076)	(0.0076)
(Log) GDP Origin			-0.4421*	-0.5259**	0.2451
			(0.2531)	(0.2460)	(0.4300)
(Log) CL Origin				0.2501	-0.6094
				(0.2899)	(0.5909)
(Log) CL GDP Origin					-0.9464**
					(0.4773)
Constant	1.1966	1.3936	4.7503**	5.6243***	4.8734**
	(1.1439)	(1.1365)	(2.1025)	(2.1760)	(2.1884)
Observations	1,095	1,095	1,095	1,095	1,095
Number of countries	154	154	154	154	154
R-squared	0.1990	0.2030	0.2314	0.2340	0.2501

Country-clustered robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$
All regressions include year dummies.

Table 3.11: Parameter estimates corresponding to Table 3.7

	(1)	(2)	(3)	(4)	(5)
(Log) WBLI	-0.0373	-0.1004	-0.0788	-0.0419	0.1488
	(0.2576)	(0.2676)	(0.2665)	(0.2753)	(0.3004)
(Log) Isolation	0.3956***	0.3550***	0.7112	2.8289	2.4772
	(0.1188)	(0.1181)	(0.4421)	(2.8772)	(2.9473)
Conflict Level 1	-0.0162	-0.0215	-0.0222	-0.0208	-0.0189
	(0.0182)	(0.0183)	(0.0182)	(0.0182)	(0.0181)
Conflict Level 2	-0.0302	-0.0307	-0.0332	-0.0295	-0.0263
	(0.0543)	(0.0556)	(0.0565)	(0.0567)	(0.0565)
(Log) GDPpc Destination	-0.0292	-0.0497	-0.0492	-0.0237	0.0642
	(0.1084)	(0.1143)	(0.1126)	(0.1168)	(0.1266)
(Log) WBLI * (Log) GDPpc	0.0170	0.0209	0.0189	0.0133	-0.0111
	(0.0262)	(0.0275)	(0.0273)	(0.0284)	(0.0324)
Contiguous Countries Conflict		-0.0224***	-0.0232***	-0.0221***	-0.0201***
		(0.0072)	(0.0071)	(0.0073)	(0.0068)
(Log) GDP Origin			-0.2383	-0.1737	0.3075
			(0.2840)	(0.2414)	(0.4768)
(Log) WBLI Origin				-1.7363	-1.0264
				(2.1102)	(2.3581)
(Log) WBLI GDP Origin					-0.7851
					(0.5949)
Constant	-1.7960	-1.2934	0.3870	4.4878	6.9055
	(1.1095)	(1.1673)	(2.2278)	(6.5542)	(5.8759)
Observations	968	968	968	968	968
Number of countries	151	151	151	151	151
R-squared	0.1826	0.1883	0.2024	0.2197	0.2253

Country-clustered robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
All regressions include year dummies.

Part IV

Final conclusions

This thesis studies the push and pull factors of internal and external forced migration comprising three essays on: (i) the determinants of dyadic refugee flows, (ii) spatial factors in internal displacement and forced migration and (iii) gender bias in refugee destinations. The first chapter introduces a gravity model, a workbench used to model international trade and migration, in order to study dyadic refugee flows for the 1990-2013 period. Two estimation methods are used so as to analyze the determinants of refugee flows: the OLS logarithmic specification and the Poisson exponential mean specification. The evidence reported has shown that armed conflicts and low standards of civil liberties lead to refugees in origin countries, while population, proximity and high standards of civil liberties attract refugees to destination countries. From a policy point of view, refugee flows could be reduced by avoiding armed conflicts and improving the standards of civil liberties in origin countries. The second chapter develops a spatial model of internal and external forced migration to study refugee and IDP counts for the period 1995-2016. The spatial model we develop represents armed conflict as a shock that takes place at a particular location and generates a migration flow. Along with the intensity of armed conflicts, we take into account geographical factors such as country size, orography and distance to other countries. The empirical evidence indicates that area and ruggedness reduce the proportion of refugees and IDPs, while the lack of civil liberties, proximity and armed conflicts lead to more refugees and IDPs. However, the intensity of armed conflicts is also important when distinguishing the evolution of both groups for the last decades: low intensity conflicts lead to refugees to a lesser extent than high intensity conflicts, which in addition also generate IDPs. The final chapter studies the gender bias in destination flows, for the period 2008-2016. In order to analyze the gender bias, we take into account factors such as civil liberties and the per capita GDP of destination countries, country of origin-related variables weighted by distance, and measures of women's rights. The evidence suggests that the gender bias is greater (more male refugees than female) in destination countries that are rich, women-friendly and have high civil liberty standards, isolated from origin countries that are poor and have low civil liberty standards.

Part V

Bibliography

Bibliography

Akgündüz, Y. E. and Torun, H. (2018). Two and a half million syrian refugees, skill mix and capital intensity. GLO Discussion Paper Series 186, Global Labor Organization (GLO).

Akram, S. (2013). Millennium development goals and the protection of displaced and refugee women and girls. *Laws*, 2(3):283–313.

Alfano, M., Dustmann, C., Frattini, T., and Fasani, F. (2016). Refugees and economic migrants: Facts, policies and challenges.

Alix-Garcia, J., Bartlett, A., and Saah, D. (2012). Displaced populations, humanitarian assistance and hosts: A framework for analyzing impacts on semi-urban households. *World Development*, 40(2):373 – 386.

Alix-Garcia, J., Bartlett, A., and Saah, D. (2013). The landscape of conflict: IDPs, aid and land-use change in Darfur. *Journal of Economic Geography*, 13(4):589–617.

Alix-Garcia, J. and Saah, D. (2009). The effect of refugee inflows on host communities: Evidence from tanzania. *World Bank Economic Review*, 24:148–170.

Anderson, J. E. (2011). The gravity model. *Annual Review of Economics*, 3(1):133–160.

Anderson, J. E. and van Wincoop, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, 93(1):170–192.

Aronow, P. M., Samii, C., and Assenova, V. A. (2015). Cluster–robust variance estimation for dyadic data. *Political Analysis*.

Ayata, B. and Yukseker, D. (2005). A belated awakening: national and international responses to the internal displacement of kurds in turkey. *New perspectives on Turkey*, 32:5–42.

Baban, F., Ilcan, S., and Rygiel, K. (2017). Syrian refugees in turkey: pathways to precarity, differential inclusion, and negotiated citizenship rights. *Journal of Ethnic and Migration Studies*, 43(1):41–57.

Baez, J. E. (2011). Civil wars beyond their borders: The human capital and health consequences of hosting refugees. *Journal of Development Economics*, 96(2):391 – 408.

Ball, P., Betts, W., Scheuren, F., Dudukovich, J., and Asher, J. (2002). Political killings in kosova/kosovo, march-june 1999. Technical report, American Association for the Advancement of Science.

Barthel, F. and Neumayer, E. (2015). Spatial dependence in asylum migration. *Journal of Ethnic and Migration Studies*, 41(7):1131–1151.

Beck, N. and Katz, J. N. (2001). Throwing out the baby with the bath water: A comment on green, kim, and yoon. *International Organization*, 55:487–495.

Bergstrand, J. H. and Egger, P. (2007). A knowledge-and-physical-capital model of international trade flows, foreign direct investment, and multinational enterprises. *Journal of International Economics*, 73(2):278–308.

Bilgili, O., Loschmann, C., and Siegel, M. (2017). The gender-based effects of displacement: The case of congolese refugees in rwanda.

Blundell, R., Griffith, R., and Windmeijer, F. (2002). Individual effects and dynamics in count data models. *Journal of Econometrics*, 108(1):113 – 131.

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.

Bohnet, H., Cottier, F., and Hug, S. (2018). Conflict-induced idps and the spread of conflict. *Journal of Conflict Resolution*, 62(4):691–716.

Bonewit, A. and Shreeves, R. (2016). Reception of female refugees and asylum seekers in the eu: case study germany. *Brussels: Directorate General for International Policies*.

Borjas, G. J. (2017). The wage impact of the marielitos: A reappraisal. *ILR Review*, 70(5):1077–1110.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249.

Cameron, A. C. and Miller, D. L. (2014). Robust inference for dyadic data. Technical report, University of California at Davis.

Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *ILR Review*, 43(2):245–257.

Ceritoglu, E., Yunculer, H. B. G., Torun, H., and Tumen, S. (2017). The impact of syrian refugees on natives' labor market outcomes in turkey: evidence from a quasi-experimental design. *IZA Journal of Labor Policy*, 6(1):5.

Crisp, J. (2010). Forced displacement in africa: Dimensions, difficulties, and policy directions. *Refugee Survey Quarterly*, 29(3):1–27.

Czaika, M. (2009a). Asylum cooperation among asymmetric countries: The case of the european union. *European Union Politics*, 10(1):89–113.

Czaika, M. (2009b). *The Political Economy of Refugee Migration and Foreign Aid*. Palgrave Macmillan.

Czaika, M. and Kis-Katos, K. (2009). Civil conflict and displacement: Village-level determinants of forced migration in aceh. *Journal of Peace Research*, 46(3):399–418.

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.

Depetris-Chauvin, E. and Santos, R. J. (2018). Unexpected guests: The impact of internal displacement inflows on rental prices in colombian host cities. *Journal of Development Economics*, 134:289 – 309.

Dreher, A., Fuchs, A., and Langlotz, S. (2019). The effects of foreign aid on refugee flows. *European Economic Review*, 112:127 – 147.

Echevarria, J. and Gardeazabal, J. (2016). Refugee gravitation. *Public Choice*, 169(3-4):269–292.

Engel, S. and Ibáñez, A. M. (2007). Displacement due to violence in colombia: A household level analysis. *Economic Development and Cultural Change*, 55(2):335–365.

Erikson, R. S., Pinto, P. M., and Rader, K. T. (2014). Dyadic analysis in international relations: A cautionary tale. *Political Analysis*.

Esen, O. and Oğuş Binatlı, A. (2017). The impact of syrian refugees on the turkish economy: Regional labour market effects. *Social Sciences*, 6(4).

Fafchamps, M. and Gubert, F. (2007). Risk sharing and network formation. *American Economic Review*, 97(2):75–79.

Fernandes, W. (2017). Internally displaced persons and northeast india. *International Studies*, 50(4):287–305.

Fisk, K. (2019). Camp settlement and communal conflict in sub-saharan africa. *Journal of Peace Research*, 56(1):58–72.

Flowerdew, R. and Aitkin, M. (1982). A method of fitting the gravity model based on the poisson distribution*. *Journal of Regional Science*, 22(2):191–202.

Goswami, U. (2007). Internal displacement migration and policy in northeastern india.

Green, D. P., Kim, S. Y., and Yoon, D. H. (2001). Dirty pool. *International Organization*, 55:441–468.

Groen, D. (2016). Simulating refugee movements: Where would you go?

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.

Holzer, T., Schneider, G., and Widmer, T. (2000). The impact of legislative deterrence measures on the number of asylum applications in switzerland (1986-1995). *The International Migration Review*, 34(4):1182–1216.

Hotelling, H. (1929). Stability in competition. *The Economic Journal*, 39(153):41–57.

Ibáñez, A. M. and Vélez, C. E. (2008). Civil conflict and forced migration: The micro determinants and welfare losses of displacement in colombia. *World Development*, 36(4):659 – 676.

Internal Displacement Monitoring Centre (2018). Global report on internal displacement.

Iqbal, Z. (2007). The geo-politics of forced migration in africa, 1992–2001. *Conflict Management and Peace Science*, 24(2):105–119.

Jones, M. (2009). Refugee status determination: three challenges. *Forced Migration Review*, (32):53–54.

Kälin, W. (2008). Guiding principles on internal displacement. *Stud. Transnat'l Legal Pol'y*, 38:1.

Keygnaert, I., Dias, S. F., Degomme, O., Devillé, W., Kennedy, P., Kováts, A., De Meyer, S., Vettenburg, N., Roelens, K., and Temmerman, M. (2014). Sexual and gender-based violence in the european asylum and reception sector: a perpetuum mobile? *The European Journal of Public Health*, 25(1):90–96.

Kondylis, F. (2008). Agricultural outputs and conflict displacement: Evidence from a policy intervention in rwanda. *Economic Development and Cultural Change*, 57(1):31–66.

Kondylis, F. (2010). Conflict displacement and labor market outcomes in post-war bosnia and herzegovina. *Journal of Development Economics*, 93(2):235 – 248.

Lanjouw, S., Mortimer, G., and Bamforth, V. (2000). Internal displacement in burma. *Disasters*, 24(3):228–239.

Leamer, E. and Levinsohn, J. (1995). International trade theory: The evidence. In Grossman, G. M. and Rogoff, K., editors, *Handbook of International Economics*, volume 3, chapter 26, pages 1339–1394. Elsevier, 1 edition.

Lischer, S. K. (2008). Security and displacement in iraq: responding to the forced migration crisis. *International Security*, 33(2):95–119.

Ljungdell, S. (1989). *Refugees? Female Asylum-seekers and Refugee Status: Guidelines for Assistance*. University of Uppsala.

Marbach, M. (2018). On imputing unhcr data. *Research & Politics*, 5(4):2053168018803239.

Martin-Storey, A., Paquette, G., Bergeron, M., Dion, J., Daigneault, I., Hébert, M., and Ricci, S. (2018). Sexual violence on campus: differences across gender and sexual minority status. *Journal of Adolescent Health*, 62(6):701–707.

McEniry, M., Samper-Ternent, R., and Cano-Gutierrez, C. (2019). Displacement due to armed conflict and violence in childhood and adulthood and its effects on older adult health: The case of the middle-income country of colombia. *SSM - Population Health*, 7:100369.

Milton, D., Spencer, M., and Findley, M. (2013). Radicalism of the hopeless: Refugee flows and transnational terrorism. *International Interactions*, 39(5):621–645.

Moore, W. and Shellman, S. (2007a). Where will they go?: a global study of refugees destinations 1965-1995. *International Studies Quarterly*, 51(4):811–834.

Moore, W. H. and Shellman, S. M. (2007b). Whither will they go? a global study of refugees' destinations, 1965–1995. *International Studies Quarterly*, 51(4):811–834.

Morales, J. S. (2018). The impact of internal displacement on destination communities: Evidence from the colombian conflict. *Journal of Development Economics*, 131:132 – 150.

Muggah, R. (2003). A tale of two solitudes: Comparing conflict and development-induced internal displacement and involuntary resettlement. *International Migration*, 41(5):5–31.

Munive, J. (2019). Resilience in Displacement and the Protection of Civilians in South Sudan. *Journal of Refugee Studies*, pages 1–21.

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. (2005a). Asylum recognition rates in western europe: Their determinants, variation, and lack of convergence. *Journal of Conflict Resolution*, 49(1):43–66.

Neumayer, E. (2005b). Bogus refugees? the determinants of asylum migration to western europe. *International Studies Quarterly*, 49(3):389–410.

Nunn, N. and Puga, D. (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94(1):20–36.

Peri, G. and Yasenov, V. (2019). The labor market effects of a refugee wave:synthetic control method meets the mariel boatlift. *Journal of Human Resources*.

Powell, R., Lawrence, A., Mwangi-Powell, F. N., and Morison, L. (2002). Female genital mutilation, asylum seekers and refugees: the need for an integrated uk policy agenda. *Forced migration review*, 14:35.

Riley, S., Degloria, S., and Elliot, S. (1999). A terrain ruggedness index that quantifies topographic heterogeneity. *Internation Journal of Science*, 5:23–27.

Roper, S. D. and Barria, L. A. (2010). Burden sharing in the funding of the unhcr: Refugee protection as an impure public good. *Journal of Conflict Resolution*.

Rüegger, S. (2019). Refugees, ethnic power relations, and civil conflict in the country of asylum. *Journal of Peace Research*, 56(1):42–57.

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(02):335–366.

Sandler, T. and Hartley, K. (2001). Economics of alliances: The lessons for collective action. *Journal of Economic Literature*, 39(3):869–896.

Sansonetti, S. (2016). *Female Refugees and Asylum Seekers: The Issue of Integration: Study*. EUR-OP.

Santos Silva, J. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Schaeffer, P. (2010). Refugees: on the economics of political migration. *International Migration*, 48(1):1–22.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

Stone, S. F. and Bang, N. J. (1999). Gravity-model specification for foreign direct investment: A case of the asia-pacific economies. *The Journal of Business and Economic Studies*, 5(1):33–42.

Suleimenova, D., Bell, D., and Groen, D. (2017). A generalized simulation development approach for predicting refugee destinations. *Scientific reports*, 7(1):13377.

Thielemann, E. R. (2004). Why asylum policy harmonization undermines refugee burden-sharing. *Eur. J. Migration & L.*, 6:47.

Tinbergen, J. (1962). *Shaping the World Economy; Suggestions for an International Economic Policy*. Twentieth Century Fund.

Torres, R. M. (2018). A crisis of rights and responsibility: feminist geopolitical perspectives on latin american refugees and migrants. *Gender, Place & Culture*, 25(1):13–36.

Tumen, S. (2016). The economic impact of syrian refugees on host countries: Quasi-experimental evidence from turkey. *American Economic Review*, 106(5):456–60.

UNHCR (2003). Sexual and gender-based violence against refugees, returnees and internally displaced persons: Guidelines for prevention and response.

UNHCR (2006). Oau convention governing the specific aspects of refugee problems in africa.

UNHCR (2011). United nations high comisioner for the refugees: Convention and protocol relating to the status of refugees.

UNHCR (2016). Age, gender and diversity accountability report 2016.

Uzonyi, G. (2015). Refugee flows and state contributions to post-cold war un peacekeeping missions. *Journal of Peace Research*.

Vanderkamp, J. (1977). The gravity model and migration behaviour: An economic interpretation. *Journal of Economic Studies*, 4(2):89–102.

Verme, Paolome, P. and Schuettler, K. (2019). The impact of forced displacement on host communities : A review of the empirical literature in economics. Technical report, World Bank.