

# No Undifferentiated Mass: Introducing a Quantitative Data Set on the Ethnic Composition of Refugee Movements\*

Seraina Rüegger<sup>†</sup> and Heidrun Bohnet<sup>‡</sup>

September 15, 2012

*Draft*

## Abstract

Refugee flows between two countries are often not homogeneous but composed of various nationalities and ethnicities. Since many refugees are caused by civil wars fought about ethno-nationalist incompatibilities and, consequently, many people are forced to flee after persecution or violence due to ethnic group membership, we argue that a disaggregated approach towards refugees that considers ethnicity is crucial. While qualitative case studies account for sub-national characteristics of refugees, quantitative comparative studies only focus on the country-level because no data on the ethnicity of refugees until now has been globally available. With our new constructed data set on the ethnic composition of refugees movements from 1975 to 2009, we want to fill this gap. After describing the data and presenting its relevance for scientist as well as policy makers, we test with a “push and pull model” the widespread assumption that refugees move along transnational ethnic linkages. The results confirm that many ethnic refugee groups flee to kin groups in neighboring countries suggesting that sub-national refugee characteristics such as ethnicity are essential in order to understand the direction of refugee movements, their diffusion effect and the refugees’ relation to the host community.

---

\*Paper prepared for presentation at ENCoRe Meeting in Oslo September 21-22, 2012. We thank Vera Troeger for helpful comments. This research was funded by the Swiss Network for International Studies.

<sup>†</sup>Corresponding Author. Center for Comparative and International Studies, ETH Zürich, Switzerland. E-mail: ruegger@icr.gess.ethz.ch.

<sup>‡</sup>University of Geneva, Switzerland. E-mail: heidrun.bohnet@unige.ch.

## 1 Introduction

More than 15 million people were refugees at the end of 2011 (UNHCR 2012). They were forced to flee because of so called "push factors": persecution, violence or civil war, finding refuge mostly in neighbouring countries. Their direction of flight or which neighboring country they choose, however, is not random. In other words, refugees are also "pulled" by specific factors into a specific host country. While there is a common agreement in the refugee literature that cultural and ethnic ties play a very important role, it has never been statistically tested. In addition, the focus within the refugee literature has been so far mainly restricted to push, or economic and political pull factors, and left out the extent of cultural pull factors. The reason could be the data shortage in this area. Until now, there exists no systematic data set on the ethnic composition of refugees. Refugees, in general, have been often regarded only as "one undifferentiated mass" (Lischer 2007, 143). It is often forgotten that they are composed of different ethnicities and, thus, can have different effects. Although several case studies have demonstrated that the refugees' ethnicity is important to determine the direction of flight, no attempt has been made yet to systematically collect data on the ethnicity of refugees. This is surprising as the refugee ethnicity is, just to name some examples, determinant for refugee status, crucial in the planning of refugee settlements to prevent ethnic rivalries and important to achieve successful local integration, as shown by researchers and policy-makers alike, such as in Kenya or Israel.

Although recent years saw an increase in quantitative refugee studies, focusing either on the push and pull factors causing refugee outflows (Davenport, Moore & Poe 2003, Moore & Shellman 2004, Iqbal & Zorn 2007) or the consequences of a refugee influx in host countries (Salehyan & Gleditsch 2006), qualitative case analysis is still predominant within the refugee-field. We argue, that a stronger focus on comparative quantitative refugee information is crucial in order to understand the causes and consequences of refugee movements as a global phenomenon and make general conclusions. Due to lack of quantitative refugee data, previous studies have several limitations as empiric refugee data provided by UNHCR or USCRI do either focus on the country of asylum or the country of origin of the refugees, but do not offer more disaggregated information on the sub-national level. However, it is very important to note that refugees are composed of different nationalities and ethnicities and because of that can also have different effects on host countries and differ in their direction of flight. For example, the Kakuma refugee camp in the north of Kenya shows high ethnic volatility, being home to ten nationalities such as Sudanese and Somali, and to distinct ethnicities of common nationality, for instance Sudanese Dinka and Didinga (Pini 2008). It is, hence, surprising that no data on

the ethnicity of refugees has yet been created.

The need of policy-makers as well as scientists for more detailed information on the ethnic background of refugees was stated by Hovy (2000, 4)<sup>1</sup>:

One of the main data limitations during the Kosovo crisis was origin. As most asylum countries record only the nationality (country of citizenship) of the applicant, few countries were able to distinguish Kosovar asylum-seekers from other citizens of the Federal Republic of Yugoslavia (FRY). ... information ... on the "ethnic origin" ... would in fact have been required. ... the importance of sub-national information on origin is likely to increase.

With a newly collected global data set on the ethnic background of refugees, we want to fill this gap and contribute to the emerging quantitative literature in refugee research. Further, future studies with more detailed refugee data will help to generate new policy implications of how to deal with large refugee movements such as those, currently, from Syria or Mali.

The paper is organized as follows: First, we review the existing refugee data and its limitations, whereas we particularly focus on data from UNHCR. Second, we briefly outline why we consider the concept of ethnicity and the sub-state focus so important. Third, we present our data set on the ethnicity of refugees, the coding instructions and preliminary summary statistics. Finally, we apply our data on an analysis of the widespread, but never tested, assumption that refugees often move along cultural linkages and flee to countries where they have ethnic kin relations. Using a push and pull approach including tempo-spatial features, we find strong empirical support that refugees indeed flee along transnational ethnic linkages, but that these are also conditional on the political performance of the country of asylum as well as the distance between host and home country and previous refugee movements.

## 2 State of the Art: Refugee Data

Forced displacement occurs in almost all regions of the world and is a global phenomenon, therefore, quantitative data with global coverage is needed to address and analyze the refugee phenomenon in a comparative manner. The United Nations High Commissioner for Refugees (UNHCR) in this regard is the main organization that offers a broad range of empiric refugee data such as annual data on refugee hosting countries or data on refugee sending countries, as well as dyadic data on refugee movements between two countries with almost global coverage

---

<sup>1</sup>Bela Hovy, Head, Population Data Unit, United Nations High Commissioner for Refugees (UNHCR), Geneva.

since the 1960s to date (UNHCR 2010). According to the UN Convention Relating to the Status of Refugees a refugee is

“a person 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 or, owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having a nationality and being outside the country of his former habitual residence as a result of such events, is unable or, owing to such fear, is unwilling to return to it” (UNHCR 2007, 17).

UNHCR counts refugees that fall under this definition, but also includes other groups of persons that are in similar circumstances as refugees such as asylum seekers, returnees, and to some part also internally displaced people (IDPs) in their data referring to them together as “persons of concern”. Their data consist mainly of stock numbers as well as limited data on the number of new arrivals. In general, the data mainly is based on the country level. UNHCR provides demographic data like sex and age, as well as data on the numbers of refugees per settlement; nevertheless, they do not provide any data on the ethnicity of refugees and, consequently are very limited in providing disaggregated refugee data. We understand that the ethnicity of refugees can be a very sensitive issue, particularly, if refugees have fled because of their ethnicity and are, thus, not willing to provide this information. However, we ensure that the data presented in the following stays anonymous and will only be used in an aggregated manner. In the end, providing information on ethnicity could help refugees to be located in a safe area where they will not be harassed by a rival ethnic group.

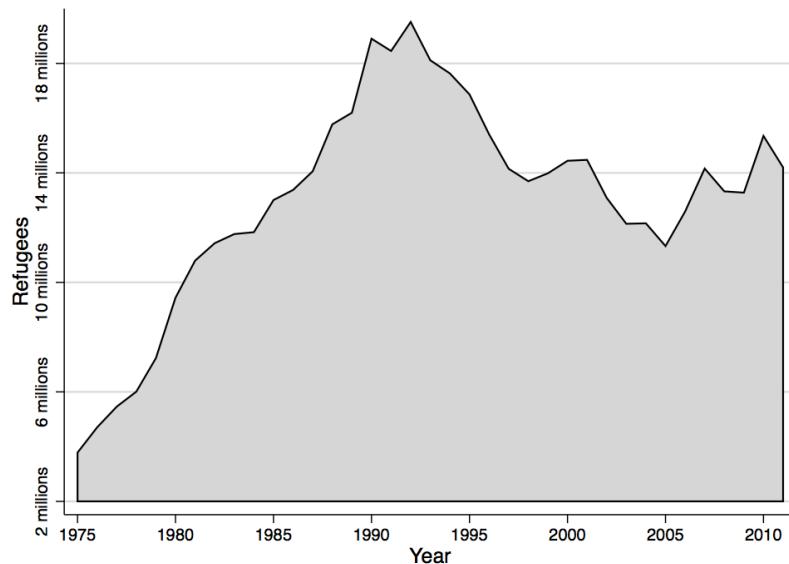
The United States Committee for Refugees (USCRI) also provides data on refugee numbers in countries of asylum or countries of origin, but also does not systematically collect data on the ethnicity of refugees.

Information on Palestinian refugees can be obtained from the United Nations Relief and Works Agency for Palestine Refugees (UNRWA 2010).

Although, all of these agencies have some information on the ethnicities of refugees in their annual reports, they do not systematically collect data on this. However, the ethnicity of refugees can be crucial for refugee protection and help to better understand, in general, refugee movements and their effects.

Graph 1 illustrates the number of refugees worldwide between 1975 and 2011. Towards the end of the Cold War, a clear peak can be observed during the early 1990s. Since then, refugee numbers slightly decreased but remain on a high level, currently above 14 millions. Thus, a considerable amount of the world’s population is directly or indirectly affected by forced migration.

**Figure 1:** Number of refugees worldwide (UNHCR and UNRWA data)



### 3 The Relevance of Ethnicity

Several authors state that ethnic civil wars or genocide, where groups are targeted based on cultural, religious or ethnic characteristics are the primary cause of refugee movements (Weiner 1996, Schmeidl 1997, Iqbal & Zorn 2007) and consequently, we argue that it is very important to assess the refugees' ethnicity. However, so far, all quantitative forced migrant studies have neglected the ethnic group membership of refugees and thus do not meet the complexity of most refugee situations. In order to introduce our data set on the ethnicity of refugees in the following section, we will now provide a definition of the ethnicity concept and explain its relevance.

We define an ethnic group as a self-perceived community with a shared culture and a common ancestry, based on Cederman (2010, 3) who defines an ethnic group as "a cultural community based on a common belief in a putative descent". Ethnic group membership can be based on a common language, religion or somatic features, but our definition does not include tribes or clans. This conceptualization of "ethnicity" complies with the Ethnic Power Relations (EPR) data set (Cederman, Wimmer & Min 2010) that identifies all politically relevant ethnic groups in a country and records the level of access to state power by their representatives.

According to the constructivist view of ethnicity, interaction unifies people as ethnic groups (Bayar 2009, 1643). Civil conflicts and wars increase interaction among people belonging to the same targeted group and consequently, facilitate the endogenous development of ethnic group awareness (Vorrath & Krebs 2009, 1). The

group-based persecution and the experience of flight strengthen the ethnic identity of a refugee group (Lischer 2005, 22 - 25) and create strong social and politicized units among them (Lebson 2010, 12). Thus, we assume that refugees are strongly aware of ethnic group membership, particularly, when fleeing from ethnic conflict or ethnic persecution.

Civil wars are the major determinant of forced migration today (Melander & Öberg 2006, 130 - 131) and approximately 52 percent of the civil wars and conflicts fought between 1946 and 2009 were ethnic conflicts<sup>2</sup>, which are identified by Weiner (1996, 29) as the primary cause of refugee flows. Thus, most refugees flee because of ethnic persecution or violence and, therefore, their ethnic background matters. Consequently, we argue that the ethnicity of refugees is a crucial element to consider in order to be able to understand refugee movements.

## 4 Ethnicity of Refugee Data Set

The refugees' ethnicity matters in two ways: First, ethnic group membership might determine whether a person becomes a forced migrant or not. This applies, particularly, to ethnic conflicts, and second, it could determine how a refugee is received in the asylum country, as, for instance, cultural similarities with the host population facilitate integration (Newland 1993, Jacobsen 1996). Hence, we argue that refugee movements should not only be analyzed on the country-level, but as well on a more disaggregated sub-national ethnic group-level, because refugees are "not one undifferentiated mass" (Lischer 2007). To fill this empirical and theoretical gap in refugee studies, we created a data set on the ethnicity of refugees which we present in the following section<sup>3</sup>.

Dyadic information on refugee movements between countries of asylum and countries of origin, as well as the sizes of the refugee movements were obtained from UNHCR (2010) and from UNRWA (2010). Refugee groups are often not homogeneous but rather consist of several ethnic groups. Thus, within each refugee movement, we systematically tried to identify up to the three largest ethnic groups and indicated their share of the total refugee flow. Many countries receive refugee flows from several countries that are not ethnically homogeneous but consist of several ethnic groups. For instance, Syria currently hosts refugees from Iraq, Somalia, and Afghanistan, as well as Palestinian refugees. Consequently, an ethnic group might

---

<sup>2</sup>A civil conflict is defined as ethnic when at least one party (the government or rebels and army factions) has ethno-nationalist claims such as self-determination, autonomy, obtaining cultural rights or ending ethnic discrimination (Wimmer, Cederman & Min 2009, 326).

<sup>3</sup>The data was collected by Seraina Rüegger, Heidrun Bohnet and Nadja Schloss under the SNIS funded project "Refugee Flows and Transnational Ethnic Linkages", a co-operation between ETH Zürich and the University of Geneva.

have kin refugees from more than one country, like the Kurds in Iraq that received kin refugees from Turkey and Iran. As a matter of feasibility, we collected the ethnicity data only for refugee flows that consisted of at least 2'000 refugees per year between neighboring countries or countries within close proximity of each other, that is a maximal distance of 950 kilometers between its borders (information on borders obtained by CShapes: Weidmann, Kuse & Gleditsch (2010)). Within this framework, we are able to provide worldwide information about the ethnic background of refugees for the years 1975 to 2009<sup>4</sup>.

Although information on the ethnicity of refugees is not provided directly by UNHCR as they do not collect data on the ethnicity of refugees systematically, it was, nevertheless, possible to collect information on the ethnicity of refugees relying on reports from UNHCR, USCRI, several NGOs, conflict narratives and news articles. We used the group list of the Ethnic Power Relations (EPR-ETH) data set (Cederman, Min & Wimmer 2008) as a source to identify ethnic groups living in a refugee sending country. In some cases refugees belonged to an ethnic group that is considered politically irrelevant and is, thus, not recorded in the EPR-ETH data set, nevertheless, we also included these groups in our refugee data set.

Precise numbers of refugees from each ethnic group were practically unavailable. But organizations working with refugees, like UNHCR, USCRI or national Refugee Councils, often indicate approximate numbers, estimations or at least evidence like “more than 31,000 [people] from Afghanistan, *mostly Hindus*, fled to India during the rise of the Taliban in the 1990s” (USCRI 2009), so that we were able to at least indicate whether a certain ethnic group within a refugee flow was dominant, a majority or a minority.

Table 1 displays the most important variables collected in our data set on three examples:

**Table 1:** Example of ECORF data set

Country of asylum	Country of origin	Year	Number of refugees	Largest ethnic group	Share of largest ethnic group	2nd largest ethnic group	Share of 2nd largest ethnic group	3rd largest ethnic group	Share of 3rd largest ethnic group
Macedonia	Serbia	1999	21'000	Albanians	DOM	Roma	MIN	Serbs	MIN
Syria	Iraq	2006	700'000	Sunni Arabs	MAJ	Shiia Arabs	MIN	Christians	MIN
Ethiopia	Eritrea	2008	21'018	Tigrinya	MAJ	Afar	MIN	Kunama	MIN

<sup>4</sup>Countries that became independent after the beginning of our sample, for instance the former Soviet countries, appear in the data set according to their independence. Further, UNHCR, for instance, treats Namibia as independent country before 1990, but we added Namibian refugees before 1990 to South Africa. Furthermore, UNHCR distinguishes between Tibetan and Chinese refugees, but we counted them all as being from China. A complete list of all coding decisions is available from the authors upon request

Further, coders were asked to specify the source of their coding, where they obtained the information on the refugees' ethnic background, and to indicate on a three-point scale their security about the coding, so that cases with lacking sources or insecurity can be excluded from analyses if required. Also, we provide the countryid of each identified politically relevant ethnic refugee group. This allows to merge the ECORF data with other data sets of interest, for instance conflict data or the Ethnic Power Relations (EPR) data set in order to obtain the access to power or the size of the refugee group in the country of origin. The original data set is shaped in country of origin - country of asylum dyadic form, but it can easily be reshaped to a country of origin or country of asylum focus as well as to the ethnic refugee group-level.

The ratio of refugees belonging to a certain ethnic group and the entire refugee flow is only estimated as either being dominant, majority or minority as exact numbers are often not available. Based on these estimations, we then calculated the absolute size of each ethnic refugee group. Of course, bearing in mind that these absolute sizes have to be regarded with great caution since the shares are estimations only. We applied the rule that if the refugee flow consisted of one dominant ethnic group, then we multiplied the size of the refugee movement, i.e. the number obtained from UNHCR, by the factor .95, since there is confidence that at least 95 percent of the flow belong to the concerned group. If there was one majority ethnic group within the refugee flow, we multiplied it by .65. If several ethnic groups were identified within a refugee movement, whereas our coding rules allowed for maximally three ethnic groups, the multiplying factors were readjusted according to the rules displayed in table 2<sup>5</sup>. The shares were defined according to the approximately 50 cases where we have precise information of the ethnic refugee group sizes: Dominant ethnic groups comprise on average 92 percent, majority groups 59 percent and minorities 19 percent of the refugee movement<sup>6</sup>.

Finally, in some countries of asylum we were not yet able to determine the ethnicity of the refugees due to lacking information. The countries concerned are Indonesia, Vietnam, Thailand and Malaysia. Thus, we suggest to exclude those countries in analyses using the Ethnic Composition Of Refugee Flows (ECORF) data set.

Although we stress the need for a sub-national refugee-focus, we do not cover individual refugee experiences of (fear of) violence or persecution and consequent

---

<sup>5</sup>The total share is mostly below 1 in order to account for insecurity.

<sup>6</sup>We assume that the actual ethnic refugee group sizes are underestimated because the underlying UNHCR numbers are end year figures. We considered to control for conflict duration (i.e. conflicts that last more than a year produce more refugees) or the end date of the conflict (more refugees from conflicts that end at the end of the year reported in UNHCR data set), but decided against it, because this would possibly lead to biased results with only better estimated numbers of refugees caused by conflict but still underestimated sizes of refugees caused by violent persecution or threat of persecution or conflict.

**Table 2:** Share of ethnic refugee group within refugee movement

Largest group	Share in %	Second group	Share in %	Third group	Share in %
Dominant	0.95				
Dominant	0.95	Minority	0.05		
Dominant	0.9	Minority	0.05	Minority	0.05
Majority	0.65				
Majority	0.6	Minority	0.3		
Majority	0.6	Minority	0.3	Minority	0.05
Minority	0.3				
Minority	0.3	Minority	0.1		
Minority	0.3	Minority	0.1	Minority	0.05

flight. Moreover, we still analyze forced migrant movements on the macro-level as this allows a quantitative global comparison.

## 5 Descriptive Statistics of the Data

In the following section, we briefly present the descriptive statistics of our data. 3'129 country-dyad-years in the UNHCR data set qualified for our ethnicity coding, that is the refugee movement consists of minimally 2'000 refugees and the countries are contiguous or in a maximal distance of 950km from each other, wherefrom we were able to code 2'836. In other words, approximately 9 percent of the relevant country-dyad-years are missing in our data set, which is a low value given the difficulty to find reliable refugee sources.

Among the 2'836 country-dyad-years we identified 5'339 ethnic group-years, which are composed of 516 individual ethnic refugee groups from 189 countries of origin. Among the 5'197 ethnic refugee groups, of which we estimated the size, 1'895 (36 percent) belong to the category dominant, 939 (18 percent) are a majority and the remaining 2'363 (46 percent) a minority. The average ethnic refugee group-year, calculated according to the previously presented estimation rules, consists of 60'000 refugees, and the median ethnic refugee group size is 7'000, while four groups constitute more than a million people: the Hutu from Rwanda in the Democratic Republic of Congo (DRC) (1994 - 1995), the Palestinians in Jordan (1992 - 2009), the Somali in Ethiopia (1979 - 1980) and the Pashtun from Afghanistan in Pakistan (1980 - 2009).

Approximately 60 percent of the identified relevant dyadic refugee movements are ethnically homogeneous, that is they are composed of one dominant ethnic group with only small additional minority ethnic groups. The remaining 40 percent are either composed of a majority with one or two minorities or several minority ethnic groups. Thus, not pursuing a disaggregated approach towards refugees would

clearly underestimate the ethnic complexity of refugee movements.

Thanks to the compatibility of our data with the EPR data set, we are able to analyze the political power status the refugees had in their home country. 3'955 (76 percent) ethnic refugee groups were politically relevant in their country of origin, among those 57 percent were excluded, had no access to state decision making, and 43 percent were included in the government at home. Among the politically excluded groups, 1'066 groups, 27 percent, were discriminated supporting the assumption that political discrimination by oppressive regimes is an important cause of refugee movements (Schmeidl 1997, 287). However, the relevance of the refugees' political status at home should not be overestimated since refugees have crossed country borders due to violence or persecution and thus did not enjoy an appropriate political situation before their flight irrespective of their ethnic group's access to state power.

## 6 Application: Do Refugees Flee Along Ethnic Linkages?

Thanks to the new data set on the ethnicity of refugees, we are able to test the widespread assumption that refugees often flee along transnational ethnic linkages to countries with fellow group members (Weiner 1996, Newland 1993).

From the total 5'339 ethnic refugee group-years identified, approximately 25 percent, 1310, fled to a country with politically relevant ethnic kin. Focusing on the absolute refugee numbers, the refugees that fled along transnational ethnic linkages make up even more: 46 percent. Furthermore, 80 percent of the refugees between 1975 and 2009 fled to a neighboring or a country in proximity of their home state and among those, 58 percent found ethnic kin in the country of asylum. Thus, although many refugees do not have kin in their receiving state, those that do flee along border-crossing ethnic linkages constitute a considerable share. In order to assess those descriptive findings statistically, we conduct a directed-dyad focused zero-inflated negative binomial regression analysis including all directed dyad-years in which the origin country of the dyad (i.e. side a) produced refugees presenting a first application of our data.

### 6.1 Literature Review

Refugees are forced to involuntarily live outside their home country and are, thus, caused by so called *push* factors. The most important push factors identified as producing refugees are political change (Weiner 1993, 95) like regime collapse (Melander & Öberg 2006, 144) or reorganizations of political communities (Zolberg, Suhrke & Aguayo 1986, 153; Davenport, Moore & Poe 2003, 43, Keely 1996,

1052). Furthermore, lack of political freedom and oppressive regimes are major sources of refugee emigration (Schmeidl 1997, 287; Moore & Shellman 2004, 729; Stanley 1987, 133) and, first of all, civil conflicts and wars (Weiner 1996, 6; Rubin & Moore 2007, 101; Martineau 2010, 145) and political violence occurring during ethnic, religious or tribal wars or genocides (Schmeidl 1997, 302; Wood 1994, 611; Newland 1993, 85-86).

Besides the stressed notion that refugees are forced to leave their country of origin and are, thus, "pushed out" of their home, several studies analyzed the direction of refugee movements and found that they are not random: refugees traditionally flee along ethnic or colonial ties into neighboring countries (Schmeidl 1997, 295), because of existing networks and low assimilation costs (Newland 1993, 86; Moore & Shellman 2007, 818). But none of these studies provides empirical support for the stated assumptions. The assumption that refugees make choices in the direction of their flight indicates that not only push factors, making the people leave their home country, but also *pull* factors, drawing the refugees in a certain direction, have to be included. An exclusive focus on push factors is too short-sighted and pull factors must be considered as well, as has been argued by Davenport, Moore & Poe (2003). Iqbal & Zorn (2007) analyze the probability of forced migration between country-dyads considering push and pull factors but neither include cultural connections nor account for the size of the refugee movement and they focus on African refugees only. Melander & Öberg (2007), on the other hand, focus on refugee numbers but consider push factors only. Thus, there is a huge gap on pull factors in the refugee literature which we will contribute to fill with our new data set.

We distinguish among three major pull categories: cultural, economic and political pull factors and in the following analysis, we focus on cultural pull factors. First, refugees fleeing along ethnic linkages because of existing networks and expected facilitated integration in countries with cultural similarities are culturally pulled. Thereby network theories become important. Refugees differ from other migrants because they involuntarily left their country without an initial motivation, thus, emigrated mainly due to the push factor violence. This leads to a stronger orientation or even political activism of the refugees towards the home country compared to other migrants who rather left their home because of pull factors like economy or family. However, pull factors such as cultural ties can also play a role for refugees. The strong distinction of refugees and other migrants is in general disputed. Refugees should also be regarded within theories of diaspora or transnationalism. Wahlbeck (2002, 234) argues that the refugees' "dual orientation both towards the society of origin and the society of settlement is not as contradictory and paradoxical as it seems" and Crisp (1999, 6) states that the process of refugee migration is affected

by transnational social networks, because migration costs are lower, if refugees can follow past refugee movement or long-term migrants from a particular country of origin (Neumayer 2004, 164).

Second, Neumayer (2004, 165) asks whether refugees are bogus and finds that richer countries in Western Europe receive a higher share of asylum applicants giving evidence that refugees are also drawn by economic pull factors. As utility-maximizers, refugees compare economic opportunities in their home country with possible countries of asylum and this affects their decision to flee (Morrison 1993, 819).

Finally, refugees are pulled by political factors. Fleeing violence and persecution or the threat thereof, refugees have incentives to relocate to places offering better security conditions, that is first, countries not experiencing conflict or war (Iqbal & Zorn 2007, 201) and second, countries with better democratic performance (Iqbal 2007, 109).

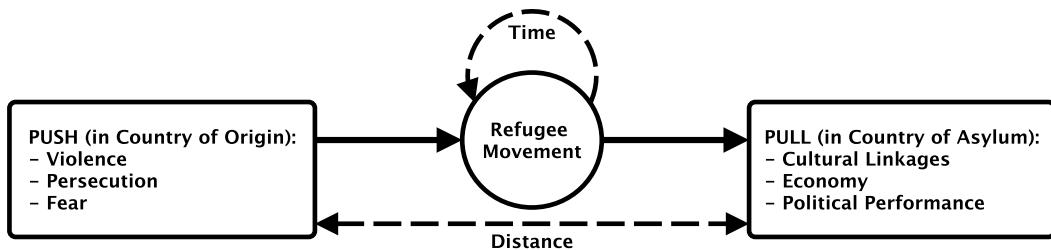
In addition, almost all authors emphasize geography, that is distance and general accessibility between country of origin and asylum, as a determinant factor for refugee flight (Iqbal 2007, Iqbal & Zorn 2007, Melander & Öberg 2007, Salehyan & Gleditsch 2006, Schmeidl 1997). Many refugees flee by foot and overland by restricted means (Schmeidl 1997, 296) and are, thus, not entirely free in their destination choice. Iqbal & Zorn (2007) found that the effect of pull factors like conflict or democracy on destination choice decreases with distance. Alike, many authors stress that time is an important factor in refugee analysis and that refugee flows are temporally dependent on earlier migration movements. While Iqbal & Zorn (2007) and Rubin & Moore (2007) find that previous refugee movements positively affect future movements, that is refugees relocate to where other refugees have gone to before, Melander & Öberg (2007, 165) argue, on the contrary, that previous refugee flows negatively impact future flows because those most willing to flee leave first generating a selection effect over time as the population remaining in the sending country will be very unwilling or unable to relocate.

Hence, we suggest that only a model including push *and* pull factors considering characteristics of the country of origin as well as the country of asylum accounts for the complexity of refugee movements, because a mere push focus is too short-sighted. This will be elaborated more detailed in the following section.

## **6.2 Tempo-Spatial Push and Pull Model to Explain the Direction of Refugee Movements**

In order to answer the question whether refugees flee along cultural lines, we propose a "push and pull" model with tempo-spatial features as illustrated in figure 2 to

analyze the direction of refugee movements, whereupon we focus on cultural pull factors in this paper, that is ethnic linkages. Since refugees share similarities with other migrants as has been suggested by Crisp (1999) and Wahlbeck (2002), a focus on push factors only would be too short-sighted, and, thus, also pull factors, like family and ethnic ties, should be taken into account when considering whereto refugees flee.



**Figure 2:** Tempo-Spatial Push and Pull Model of Refugee Flight

A transnational ethnic kin group is defined as a group that lives in more than one country, for instance the Kurds who are found in different countries: Turkey, Syria, Iraq, Iran and others. Since border-crossing groups are often caused by arbitrary border definitions from greater powers, transnational ethnic kin groups are mostly found in neighboring countries. A group with transnational kin linkages affected by conflict has several consequences: First, the group is likely to draw international attention and involvement, particularly from the trans-border kin which may cause the diffusion of conflict along these cultural linkages (Gleditsch 2007, Cederman et al. 2012). Second, the conflict-affected group is likely to seek refuge among the kin group across the border. Transnational kin groups often feature established border-crossing networks, that is transborder kin facilitates the flow of information as a consequence of shared media, personal contacts or “cognitive shortcuts” and shared values (Zhukov & Stewart Forthcoming, Simmons & Elkins 2004), so that refugees will know about the current situation in a possible host country with transnational ethnic kin. Further, cultural similarities between refugees and the host population facilitate integration and accommodation (Newland 1993, Schmeidl 1997, Moore & Shellman 2004). Assuming that refugees are pulled by cultural factors, we, therefore, hypothesize that:

*Hypothesis: Refugees move along transnational ethnic linkages.*

Although we present the above illustrated pull factors separately, we strongly assume that they are interlinked, that is, that refugees particularly follow cultural ties if the economic and political situation in the host country is better compared to the sending country.

Refugees are also pulled by political factors, that is refugees tend to flee to countries with a better political performance and to countries that do not experience conflict. First, we define a country as politically performing well if its political system complies with democratic values like basic rights, separation of powers, freedom of speech etc. and if representatives of all groups in the country have access to governmental power. Since refugee flee persecution and violence, they have comprehensible incentives to move to a country of asylum where the risk of being the target of violence again is low. Democratic countries are less likely to repress or discriminate their population or certain groups within the population. Hence, refugees are better off in democratic host countries where their basic rights are secured. Referring to refugees with transborder ethnic kin groups, the above described exchange of information along transnational networks will ensure that refugees know about the political situation of their border-crossing kin and they are unlikely to flee to countries where their kin group is excluded from the government, as this means that the kin group has no power at the national level to advocate for their kin refugees, for instance, to release financial resources for better refugee accommodation. Second, refugees will also refrain from moving to possible countries of asylum that experience conflict themselves and particularly where their kin group is involved in conflict as the chances are high of suffering from persecution and violence again or being treated equally bad by the host government as at home.

While refugees do not participate in state politics in the host country as they lack citizenship rights, the political pull argument still holds in this respect, because, first, democratic and inclusive governments tend to respect human rights which affect refugees and, second, refugees might be interested in participation possibilities in case they do not return to their country of origin and are naturalized in the future. In addition, economic pull factors affect refugees' destination choice. Similar to the above political mechanism, refugees, seen as rational utility maximizers, should flee to countries with a better economic performance than their home country because wealthier countries have more resources to accommodate refugees. Moreover, the prospects of personal well-being at a later stage of integration are better. This causal effect similarly applies to refugees fleeing along transnational ethnic networks. They are likely to consider the economic performance of their host country and particularly of their ethnic kin group, since not all groups in a country are economically in the same position. By observing the economic position of their ethnic kin group in the various possible host countries, we assume, refugees choose the one where the ethnic kin group is best off in economic terms.

Consequently, we will control for both political and economic pull factors impacting refugee movements along ethnic linkages.

Besides the push and pull factors, tempo-spatial mechanism affect refugee move-

ments. First, refugees are more likely to flee to non-distant countries because of feasibility. Second, refugees are temporally dependent on previous refugee movements because previous flows facilitate future flows thanks to established networks of transportation and information exchange. Thus, we assume that the above described pull factors are conditional on the distance between the country of origin and the country of asylum and the time, that is previous refugee movements. As we focus on pull factors, we will only analyze countries producing refugees in order to control for push factors. This will be explained in more detail after presenting the measurement of our theoretical concepts in the next chapter.

### 6.3 Operationalization and Descriptive Statistics

The *dependent variable* is the number of refugees belonging to the same ethnic group moving within a country dyad, that is the count of refugees sharing ethnic group membership from the same country of origin in the same country of asylum in each year. We use our newly compiled data set in order to identify ethnic refugee groups and their sizes<sup>7</sup>. Since conflicts and consequent refugee movements are rare events, most countries do neither produce nor host refugees. The average ethnic refugee group identified in our data set consists of approximately 60'000 refugees, the minimum are 100 refugees and the maximum 3 million refugees (ethnic Pashtun from Afghanistan in Pakistan in the late 1980s).

The main *explanatory variable* is whether the refugees are connected by ethnic ties to their possible country of asylum. Hence, we need to know if the country of origin and the possible country of asylum are linked by ethnic ties. The transnational ethnic kin (TEK) data (International Conflict Research Group, ETH Zürich 2011) enumerates all politically relevant ethnic groups that are present in more than one state. The group list is based on the Ethnic Power Relations (EPR-ETH) data set that identifies all politically relevant ethnic groups in a state and records the level of access to state power for the representatives of each (Cederman, Min & Wimmer 2008). TEK linkages are not endogenous to previous refugee movements, because trans-border ties are coded only for politically relevant ethnic groups. Migrant groups like refugees are per definition of EPR never politically relevant. The dummy variable is coded 1 for all country-dyads where the refugee group is affected by ethnic linkages and 0 for those without. Transnational ethnic linkages, indeed, affect most countries: Only 15 countries worldwide do not have major TEK linkages, for example, New Zealand, Norway, Portugal, or Sierra Leone.

---

<sup>7</sup>Since refugee numbers are end-year figures, we do not need to lag the independent variable as we are able to clearly identify the causal mechanisms that produced the refugees as a temporal function.

Further, we include several *control variables* to account for alternative explanations of refugee movements, whereas we distinguish, on the one hand, between pull and tempo-spatial factors and, on the other hand, between ethnic group- or country-related factors.

To account for group-related *pull* factors affecting the direction of refugee flows we include several variables: First, we created two dummy variables that indicate whether the TEK link of the refugees concerns a politically powerful or powerless group in the country of asylum. Information on the political status of ethnic groups was obtained through the EPR-ETH data set, whereas we distinguish between excluded and included groups. Powerless or marginalized groups (meg) have one of the following EPR statuses: powerless, discriminated, regional or separatist autonomy (Cederman, Wimmer & Min 2010, 100-101), as opposed to either monopoly, dominant, senior or junior partner groups that are coded as ethnic groups in power (egip). Second, to account for the fact that refugees rather flee to countries with a large ethnic kin population, because a bigger population can more easily absorb the inflow of people (Iqbal 2007, 108), we include the relative size of the ethnic kin group. We assume that kin refugees of a relatively large ethnic group experience a facilitated accommodation and integration. Data on group sizes is taken from EPR-ETH.

Third, we assume that refugees fleeing violence, persecution and conflict in their home country intend to avoid such situations in their asylum state and thus, only flee towards ethnic kin groups who do not experience conflict themselves. Also on the country-level, we control for ongoing conflict. Therefore, we include two dummy variables of whether the kin group and/or the country of asylum experiences conflict. In order to avoid reverse causality, since refugees are also a possible cause of conflict (Salehyan & Gleditsch 2006), we lagged these variables one year. Data on conflict incidences was obtained from UCDP/PRIO. The UCDP/PRIO Armed Conflict data set (Gleditsch et al. 2002, Version V.4\_2010) records all instances of civil conflicts and wars with a minimum of 25 battle deaths. The Uppsala Conflict Data Project defines an armed conflict “as a contested incompatibility that concerns government or territory or both where the use of armed force between two parties results in at least 25 battle-related deaths. Of these two parties, at least one is the government of a state” (Gleditsch et al. 2002, 618-619). Group-related conflict data is drawn from the Non State Actor Docking Project (Wucherpfennig et al. 2012) which identifies rebel groups with ethnic claims or recruitment along ethnic lines within the updated Expanded Uppsala Armed Conflict Data (EACD) (Cunningham, Gleditsch & Salehyan 2009) upon which ethnic conflicts are identified.

Further, we consider country-related factors: To test the assumption that refugees are pulled by the political performance of a country, we include the difference in

the share of excluded population between the host and the origin country given by EPR-ETH, because we assume that refugees move towards host countries where a high share of people is included in state politics, that is where the political performance is fair and better than in the origin country. As an alternative measurement of the political performance of a host country, we include the difference in the Polity IV index (Marshall, Jaggers & Gurr 2012) between the host and the source country. We assume that refugees flee to countries that are more democratic than their country of origin. However, a curvilinear relationship between the Polity IV value of the host country and the number of refugees could also be considered, because anocratic or weak states are less able to secure their borders and, thus, could receive more refugees (Adamson 2006).

In addition, we assume that refugees are pulled towards wealthier countries than their home country, because poor countries have less capacities to accommodate refugees, since providing shelter and food depends on financial and natural resources. Therefore, we include the difference in the annual gross domestic product (PPP adjusted real per capita GDP) between the host and the source country. GDP data are taken from several sources (Heston, Summers & Aten 2011, Fearon & Laitin 2003, World Bank 2011, Gleditsch 2008). And we include the population size (logged) of the country of asylum, because, as mentioned above, more populous countries more easily absorb a high amount of refugees. Population data are taken from Penn World Tables (Heston, Summers & Aten 2011).

In the above presented country-level control variables we use the difference between the sending and the receiving country and not only the political and economic performance of the refugee-receiving country alone, because otherwise, we would neglect the refugees' orientation towards their country of origin. Furthermore, economic and political factors often cluster spatially, meaning that countries in the same region often have similar values in global comparison and many refugees flee to neighboring countries, because of a shorter distance.

Although we focus on pull factors exclusively, we consider the push mechanism violence and control for the severity of the conflict that produced the refugees by including the number of battle deaths provided by Lacina & Gleditsch (2005) in the country of origin, because a severe conflict is likely to produce more refugees.

We consider the *spatial* dimension of refugee movements by controlling for the minimal distance between the origin and the host country (Weidmann, Kuse & Gleditsch 2010), because refugees are more likely to relocate to neighboring countries, that is when the minimal distance equals zero. Finally, to account for the *temporal* dependence we measure the years with refugee movements within a dyad. To account for duration dependence, we use natural cubic splines of years with no refugee movement with three knots (Beck, Katz & Tucker 1998).

Table 3 displays the descriptive statistics of the variables used in the following analysis of the direction of refugee flows.

**Table 3:** Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Group size	7503.906	77299.667	0	3108675.5	42517
TEK linkage	0.137	0.344	0	1	42564
TEK group egip	0.045	0.206	0	1	42564
TEK group meg	0.087	0.282	0	1	42564
TEK group size	0.028	0.124	0	1	42564
TEK group conflict (lagged)	0.006	0.077	0	1	42564
Conflict in CoA (lagged)	0.229	0.42	0	1	39463
D_Exclusion	-0.139	0.368	-0.97	0.98	39798
D_Polity 2	1.085	7.157	-20	19	33330
D_GDP	2937.758	8301.549	-23947.91	80206.883	39091
Population in CoA (ln)	9.388	1.447	5.432	14.096	40605
Battle deaths	3884.868	9121.380	0	80000	20871
Neighbors	0.294	0.455	0	1	42564
Flow years	1.006	3.837	0	35	42564

## 6.4 Analysis

In order to assess the direction of refugee movements, we analyze all possible countries of asylum of an ethnic refugee group and try to identify the factors pulling the refugees towards a certain destination. We define possible countries of asylum as all countries that are situated in proximity, that is no farther than 950km<sup>8</sup>, to the refugee's country of origin. Hence, the *unit of analysis* are directed dyad-years where the origin country in the dyad produced refugees. Since we disaggregate refugee movements to the ethnic group level, dyad-years may appear up to three times in the data set, for instance ethnic Azande and ethnic Dinka fled from Sudan to the Central African Republic simultaneously.

The main dependent variable, the number ethnic refugees, is a count variable with overdispersion and zero-inflation and consequently, we use zero-inflated negative binomial regression with country of origin-clustered robust standard errors to estimate the impact of the above described pull factors on the direction of forced migrants. The zero-inflated negative binomial regression model consists of two parts, first, a count equation estimating the number of refugees in a dyad and, second, the inflation equation where the probability of zero refugees in a dyad is estimated. The strong zero-inflated count of refugees yields from the fact that most dyads do not experience refugee movements, because either the two countries are non-contiguous

<sup>8</sup>The 950 kilometer threshold is of course arbitrary. We assume that a distance of 950km is still feasible to cross with restricted means, i.e. car, bus, train, as is often the case for refugees.

and do not have any ties like cultural linkages or established networks of exchange or because refugees often do not flee to all possible countries of asylum.

Table 4 shows the results of the regression models. The first model displays a refugee count equation for all directed-dyads with possible countries of asylum and the according logit model of the probability of no refugees. The main independent variable is whether the refugees have transnational ethnic linkages to the country of asylum. The polity and the battle deaths control variables are omitted in the second model resulting in a higher number of observations as well as non-zero observations. In the third model, the refugees' TEK linkages are disaggregated to whether they affect politically included (ethnic groups in power - egip) or excluded (marginalized ethnic groups - meg) ethnic groups.

The first model confirms our assumption that refugees move along transnational ethnic linkages: Country-dyads where the refugees produced in the origin country have transborder ethnic ties with a possible country of asylum receive significantly higher numbers of refugees than those country-dyads without ethnic linkages. Thus, pull factors matter and the hypothesis that refugees are pulled by ethnic ties is supported. However, if the TEK group is involved in conflict, less refugees are likely to move to the concerned possible country of asylum, as the coefficient is negative and significant at the 5 percent error level. In contrast, against expectations, more refugees tend to flee to countries experiencing conflict. Since we lagged the conflict variable one year, we are certain, that this conflict is not induced by the refugees. The positive and significant impact of the conflict variable might be explained by the fact, that conflicts cluster spatially, thus, that countries neighboring a refugee-producing country that is very likely affected by conflict do also likely experience conflict. Consequently, refugees are rather not pulled by conflict, but the positive impact of conflict on refugees is a spatial effect. The remaining control variables do not have a significant impact indicating that refugees do rather not consider the political or economic performance of the host country. But temporal factors have a strong impact on the count of refugees: the number of refugees within a dyad is positively and significantly affected by previous refugee movements.

The inflation equation in the first model reveals that the absence of transborder ethnic ties increases the chances of not hosting refugees. However, the probability of zero refugees is mainly determined by tempo-spatial factors: non-contiguous countries have a significantly increased chance of not hosting refugees. Neighboring dyads are more likely to observe refugee movements. Hence, distance matters and refugees due to their restricted means of transportation most likely relocate to neighboring countries. Further, the probability of receiving no refugees significantly increases with every year the dyad did not observe refugees. Thus, temporal dy-

**Table 4:** Regression Results, Number of ethnic refugees

	Model 1	Model 2	Model 3
Transnational ethnic linkage	1.212*** (0.297)	1.007*** (0.288)	
TEK group in CoA included			0.665 (0.463)
TEK group in CoA excluded			0.847*** (0.298)
TEK group size	0.486 (0.956)	0.181 (0.729)	0.464 (0.786)
TEK group conflict (lagged)	-1.240** (0.592)	-1.099** (0.464)	-1.069** (0.477)
Conflict in CoA (lagged)	0.401*** (0.149)	0.516*** (0.177)	0.479*** (0.163)
D_Exclusion	0.530* (0.283)	0.475 (0.341)	0.244 (0.377)
D_Polity 2	-0.011 (0.013)		
D_GDP	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
Population in CoA (ln)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Battle deaths in CoO	-0.000 (0.000)		
Neighbors	0.651** (0.309)	0.619 (0.420)	0.710* (0.391)
Flow years	0.097*** (0.024)	0.065*** (0.012)	0.058*** (0.014)
Constant	9.051*** (0.356)	9.216*** (0.308)	9.328*** (0.310)
<i>Inflation</i>			
Transnational ethnic linkage	-1.080*** (0.329)	-0.938*** (0.205)	-0.944*** (0.203)
Conflict in CoA (lagged)	-0.169 (0.192)	-0.155 (0.169)	-0.155 (0.169)
Neighbors	-1.804*** (0.290)	-1.796*** (0.235)	-1.796*** (0.236)
No flow years	3.002*** (0.206)	3.154*** (0.168)	3.156*** (0.168)
Spline 1	0.148*** (0.011)	0.156*** (0.008)	0.156*** (0.008)
Spline 2	-0.055*** (0.005)	-0.059*** (0.004)	-0.059*** (0.004)
Spline 3	0.007*** (0.002)	0.008*** (0.001)	0.008*** (0.001)
Constant	0.875*** (0.275)	0.721*** (0.217)	0.720*** (0.217)
Observations	16959	38722	38722
Non-zero Observations	2093	4374	4374
chi2	127.257	85.406	94.090

Standard errors in parentheses (clustered on country of origin)

\* p&lt;0.1, \*\* p&lt;0.05, \*\*\* p&lt;0.01

namics are important because refugees follow previous forced migrants.

In the second model, we excluded the control variable measuring the difference in the Polity value between the country of origin and the country of asylum and the number of battle deaths in the country of origin, because of the many missing values. Now we have an increased number of observation and also more non-zero observations, that is directed dyads experiencing a refugee movement. As in model 1, the predicted number of refugees is higher in country-dyads where the refugees have ethnic ties with the country of asylum. But the significant and negative coefficient for whether the TEK group experiences conflict again reveals that refugees are less likely to follow ethnic ties if the ethnic linkages concern a kin group that is involved in conflict. Thus, refugees tend to avoid host countries where their kin group experiences violent conflict. This is comprehensible since refugees were forced to flee due to conflict and, consequently, do not want to experience the same conditions again.

In order to test the assumption that refugees consider the political status of their kin groups, in model 3, the main explanatory variable is divided into two dummy variables that measure, first, whether the transnational ethnic kin group of the refugees is politically powerful or, second, whether it is excluded from central decision making. The hypothesis that refugees flee towards politically powerful ethnic kin groups is not supported. To the contrary, the predicted count of refugees is significantly higher if the refugees have cross-border ties to a politically marginalized group, but the coefficient for included TEK groups is insignificant. This leads to the conclusion that refugees only consider cultural similarities in their destination choice but not political arguments. Two reasons explain why the political status of transnational kin groups does not deter or pull refugees: First, much more TEK groups are politically excluded than included, so that the insignificant impact of the powerful TEK variable might stem from the lower number of observations. Second, refugees, as long as they do not naturalize, will not have political rights in the country of asylum and, thus, political opportunities play a less important role, particularly in the short-term. Further, there is a neighborhood effect because significantly more refugees move to countries contiguous to their home state. But country-level factors like the difference in the political or economic performance of the home and the possible host country or the population size do not affect the destination choice of the refugees.

The inflation models of model 2 and 3 do not differ from the first inflation. The probability of hosting refugees depends on transnational ethnic linkages, distance and particularly time, i.e. previous refugee movements.

In neither of the three models, the population size, the share of excluded population or the GDP play an important role since the coefficients are insignificant and zero.

Thus, refugee movements are not affected by a better political or economic performances of the host compared to the source country. The low impact of the GDP on refugee flow direction could be explained by the fact that refugees significantly move to neighboring or nearby countries and wealth often clusters regionally, that is poor countries often neighbor poor countries. Neither has the size of the refugees' ethnic kin a significant impact in any of the three models. An explanation for the insignificance of the refugees' ethnic kin group's size is that the relevant threshold for refugees to move towards a certain direction is already met if any TEK links exist no matter how big the concerned group is.

In all models, we tested for overdispersion of the data with the parameter  $\alpha$ . The positive and significant  $\alpha$  confirms that the observations are overdispersed, hence, we use a negative binomial and not a poisson distribution. The Vuong test conducted on the first model without robust standard errors confirms that the zero-inflated binomial regression model fits our data better than a negative binomial model, because we have so many zero observations.

## 6.5 Sensitivity Analysis

As a robustness check of the results supporting our hypothesis that refugees follow ethnic ties, we calculate three additional models displayed in table 5.

Model 4 includes dummy variables for the region of the country of origin in order to correct for unit-specific heterogeneity. They comprise Europe (baseline category), America, Subsaharan Africa, North Africa, Middle East, West Asia and South-East Asia. The partially significant coefficients for the world regions reveal that there are regional differences, but the general positive and significant impact of border-crossing ethnic connections on the predicted count of ethnic refugees does not change. Similarly, the negative and significant effect of kin groups experiencing conflict on the number of refugees withstands. In an additional model, that is not displayed in the paper due to lack of space, we included country dummies for the 69 countries of origin that produced refugees instead of the regional dummies. This, however, does neither affect our results.

The fifth model is estimated for ethnic refugee movements consisting of less than 100'000 refugees only. It excludes the observations above the 98.5th percentile and allows to control for whether the outliers with high numbers of refugees up to above a million have strong leverage on the results. The coefficient for ethnic linkages is still positive and significant, although only with a 10 percent error margin. Likewise, the effect of TEK groups that experience conflict is still negative but with slightly higher standard errors, which in both cases might be a consequence of the smaller number of observations. Again, the positive impact of conflict is, counter-

**Table 5:** Robustness Checks

	Model 4	Model 5	Model 6
Transnational ethnic linkage	0.718** (0.297)	0.350* (0.193)	0.467** (0.208)
TEK group size	0.107 (0.715)	-0.041 (0.333)	1.891*** (0.541)
TEK group conflict (lagged)	-0.829** (0.389)	-0.453* (0.239)	0.127 (0.323)
Conflict in CoA (lagged)	0.193 (0.173)	0.338** (0.132)	0.168 (0.159)
D_Exclusion	0.362 (0.260)	0.155 (0.162)	0.054 (0.184)
D_GDP	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)
Population in CoA (ln)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Neighbors	0.400 (0.264)	0.284 (0.215)	1.875*** (0.243)
Flow years	0.042*** (0.009)	0.019** (0.007)	
America	-0.581* (0.344)		
Subsahara Africa	0.217 (0.401)		
North Africa	0.572** (0.232)		
Middle East	1.292** (0.608)		
West Asia	2.005*** (0.492)		
Southeast Asia	0.499 (0.485)		
No flow years		-3.075*** (0.173)	
Spline 1		-0.152*** (0.009)	
Spline 2		0.057*** (0.004)	
Spline 3		-0.008*** (0.001)	
Constant	9.236*** (0.387)	9.008*** (0.198)	-0.887*** (0.280)
<i>Inflation</i>			
Transnational ethnic linkage	-0.929*** (0.202)	-0.780*** (0.165)	
Conflict in CoA (lagged)	-0.160 (0.168)	-0.125 (0.153)	
Neighbors	-1.781*** (0.231)	-1.696*** (0.231)	
No flow years	3.137*** (0.168)	3.017*** (0.174)	
Spline 1	0.155*** (0.009)	0.149*** (0.009)	
Spline 2	-0.058*** (0.004)	-0.056*** (0.004)	
Spline 3	0.008*** (0.001)	0.008*** (0.001)	
Constant	0.740*** (0.215)	0.792*** (0.216)	
Observations	38722	38108	38758
Non-zero Observations	4374	3760	
chi2	364.819	37.508	1259.189

Standard errors in parentheses (clustered on country of origin)

\* p&lt;0.1, \*\* p&lt;0.05, \*\*\* p&lt;0.01

intuitively, significant indicating that more refugees flee to countries experiencing conflict, even if conflict experiencing kin groups still have a negative and significant effect on refugees.

In the sixth model we estimate the probability of receiving ethnic kin refugees with a simple logit model, thus the dependent variable is binary (1 for country-dyads with refugees and 0 without). Same as in the count models, countries where refugees find ethnic kin have a higher probability of hosting those refugees. Further, again, time dynamics and space affect the chances of receiving refugees. In contrary to the previous models, on the one hand, the coefficient for the relative size of the TEK group is positive and significant. Thus, if not considering the size of the refugee movement but only whether any refugees are present or not, larger ethnic kin groups have a higher probability of hosting refugees. On the other hand, conflict experiencing TEK groups do not have a significant impact on the probability of attracting refugees. Since most countries never or seldom host refugees, we also estimated the same model with a rare events logit (King & Zeng 2001), but, besides minor changes in the standard errors, the results remain the same.

## 6.6 Results

Our analysis of the direction of refugee movements reveals three main findings: First, we find strong empirical support that refugees move along transnational ethnic connections, because countries that are linked by ethnic ties to a refugee group have a higher predicted count of refugees than those countries without. Thus, although forced to leave their country of origin due to violence or persecution, refugees are not only affected by push factors but do also consider pull factors directing them towards certain countries of asylum. Particularly, cultural pull factors affect the refugees' destination choice, because, on the one hand, cultural connections provide established networks of transportation facilitating the flight and, on the other hand, cultural commonalities with the host population facilitate integration.

Second, with the application of our new data set that disaggregates refugee movements between two countries on to the ethnic group-level, we also find strong empirical support that refugee movements along transnational ethnic ties depend on the living conditions of the refugees' ethnic kin group. Country-dyads where the refugees' kin group is involved in conflict have a lower predicted count of refugees. The causal mechanism behind is that refugees fleeing violence and persecution do not want to be confronted with the same conditions in a possible asylum country. However, our results do not support the assumption that refugees consider the political status of their kin group that is that refugees are pulled towards border-crossing kin that is politically powerful.

Finally, our models reveal a strong tempo-spatial dependence of refugee movements: Refugees are more likely to move to neighboring states and often refugees follow previous migration movements and existing networks of flight.

## 7 Conclusion

In this paper we introduce a new data set on the ethnicity of refugees. Previous quantitative refugee-research focused on the refugee-receiving or sending countries only and neglected sub-national refugee characteristics. As Lischer (2005) states refugees are not an “undifferentiated mass” but instead refugee movements are often highly heterogeneous and composed of different ethnic groups with different relations to their home and asylum country. With the new global data set on the ethnic composition of refugee flows we fill this gap and allow to conduct disaggregated analyses on refugees beyond the country-level. Our data set indicates up to the three largest ethnic refugee groups within each country-dyadic refugee movement for each year with at least 2'000 refugees from 1975 to 2009. As a matter of feasibility, we collected ethnicity data for refugee flows between contiguous and nearby countries only, but we are still able to cover more than 80 percent of the world's refugee movements.

We then proceed by presenting an application of our data with a test of the assumption that the direction of refugee flows is not random but that refugees flee along transnational ethnic linkages. Assuming that refugees are not only affected by so-called push factors, that is violence and persecution in the country of origin, but also pull factors impacting the direction of the refugee movement, we apply a push and pull approach with tempo-spatial features to analyze the direction of refugee flows. While previous studies particularly analyzed the different push factors producing refugees, only few authors focused on pull mechanism finding out that refugees tend to be pulled to democratic and wealthy states. Furthermore, earlier studies stressed that both political and economic pull factors highly depend on tempo-spatial measures, that is refugees tend to flee to less distant countries and they are subjected to strong temporal dependence. However, no study so far has focussed on cultural pull factors. Our application fills this gap. We assume, that refugees flee along ethnic ties and that this is conditional on the political and economic performance of the host country in comparison to the sending country. To test this hypothesis, we conduct zero-inflated negative binomial regression models with the number of ethnic refugees as dependent variable and directed country-dyad-years of all possible countries of asylum as units of analysis. We find strong empirical support that country-dyads with ethnic ties observe higher numbers of refugees. Moreover, refugees consider the living conditions of their kin groups and

are deterred by kin groups experiencing conflict themselves since refugees fleeing conflict and violence do not want to experience such situations in the asylum state again. However, we did not find evidence that refugees are pulled by other political or economic factors. Finally, spatio-temporal factors matter significantly: Most refugees flee to nearby or neighboring countries and previous refugees movements positively affect future flows.

As we already demonstrated, our data allows for different research designs, for instance, with a focus on ethnic refugee groups and directed country-dyads of the refugee sending and all possible countries of asylum, as in our approach. Further research on the direction of refugee flows could use country-dyads or ethnic groups as unit of analysis and assess which ethnic groups have the highest chance of hosting refugees. And of course other fields might use our data as well, for instance, the growing literature on the protection and security challenges of refugees in the country of asylum or the impact of the ethnicity of the refugees on refugee policies. Since not only the causes why refugees flee along ethnic linkages but also the consequences thereof are important to be studied.

In general, detailed information and data on refugees is crucial, for example, refugee camp statistics are necessary in order to calculate how much resources, food and general assistance is needed (Alix-Garcia & Saah 2008, 7). Further, UNHCR provides information on the age and the sex of most refugees, because such social factors determine the type and amount of aid needed. In the same way, we suggest, that knowledge about the ethnicity of refugees is important in order to help and assist the refugees according to their precise needs. For instance, putting challenging ethnic groups that might have been opposite parties in the conflict at home together in one camp could create insecurity, such as has been observed in Kenya (Crisp 1999, 9, 29). Information about the ethnic composition of refugee movements allows for more appropriate planning, to guarantee safety to refugees and prevent future rivalries between ethnic groups. Finally, data on the ethnicity of refugees could also help to better understand integration mechanism and, consequently, help to achieve successful local integration.

## References

Adamson, Fiona B. 2006. "Crossing borders: International migration and national security." *International Security* 31(1):165–199.

Alix-Garcia, Jennifer & David Saah. 2008. "The Effect of Refugee Inflows on Host Communities: Evidence from Tanzania." Online: <http://ssrn.com/abstract=836147>. Accessed: 18.6.2012.

Bayar, Murat. 2009. "Reconsidering primordialism: an alternative approach to the study of ethnicity." *Ethnic and Racial Studies* 32(9):1639–1657.

Beck, Nathaniel, Jonathan N. Katz & Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(4):1260–1288.

Cederman, Lars-Erik. 2010. Nationalism and Ethnicity in International Relations. In *The Handbook of International Relations*, ed. Walter Carlsnaes, Thomas Risse & Beth Simmons. 2 ed. Sage.

Cederman, Lars-Erik, Andreas Wimmer & Brian Min. 2010. "Why Do Ethnic Groups Rebel? New Data and Analysis." *World Politics* 62(1):87–119.

Cederman, Lars-Erik, Brian Min & Andreas Wimmer. 2008. "Ethnic Armed Conflict dataset." Online: <http://hdl.handle.net/1902.1/11797>. Accessed: 4.7.2010.

Cederman, Lars-Erik, Kristian Skrede Gleditsch, Idean Salehyan & Julian Wucherpfennig. 2012. "Transborder Ethnic Kin and Civil War." *International Organizations*. Forthcoming.

Crisp, Jeff. 1999. "Policy challenges of the new diasporas: migrant networks and their impact on asylum flows and regimes." *Policy Research Unit, UNHCR* pp. 1–11.

Cunningham, David E., Kristian Skrede Gleditsch & Idean Salehyan. 2009. "It Takes Two." *Journal of Conflict Resolution* 53(4):570–597.

Davenport, Christian A, Will H Moore & Steven C Poe. 2003. "Sometimes you just have to leave: Domestic threats and forced migration, 1964–1989." *International Interactions* 29(1):27–55.

Fearon, James D. & David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97(01):75–90.

Gleditsch, Kristian Skrede. 2007. "Transnational Dimensions of Civil War." *Journal of Peace Research* 44(3):54–66.

Gleditsch, Kristian Skrede. 2008. "Expanded GDP Data 5.0 Beta." Online: <http://privatewww.essex.ac.uk/~ksg/data/expgdpv5.0.zip>. Accessed: 5.6.2011.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg & Havard Strand. 2002. "Armed Conflict 1946–2001: A New Dataset." *Journal of Peace Research* 39(5):615–637.

Heston, Alan, Robert Summers & Bettina Aten. 2011. "Penn World Table Version 7.0, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, May 2011." Online: [http://pwt.econ.upenn.edu/php\\_site/pwt\\_index.php](http://pwt.econ.upenn.edu/php_site/pwt_index.php). Accessed: 14.3.2012.

Hovy, Bela. 2000. "'Statistically Correct' Asylum Data: Prospects and Limitations." *Paper presented at the Joint ECE-EUROSTAT work session on Migration Statistics, Geneva*.

Iqbal, Zaryab. 2007. "The Geo-Politics of Forced Migration in Africa, 1992–2001." *Conflict Management and Peace Science* 24(2):105–119.

Iqbal, Zaryab & Christopher Zorn. 2007. "Civil War and Refugees in Post-Cold War Africa." *Civil Wars* 9(2):200–213.

Jacobsen, Karen. 1996. "Factors Influencing the Policy Responses of Host Governments to Mass Refugee Influxes." *International Migration Review* 30(39):655–678.

Keely, Charles B. 1996. "How Nation-States Create and Respond to Refugee Flows." *International Migration Review* 30(4):1046–1066.

King, Gary & Langche Zeng. 2001. "Explaining Rare Events in International Relations." *International Organization* 55(03):693–715.

Lacina, Bethany & Nils Petter Gleditsch. 2005. "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths." *European Journal of Population* 21(2-3):116–145.

Lebson, Mike. 2010. "Why Refugees Rebel: The Politics of Refugee Militarization." *Paper presented at the Annual Conference of the International Studies Association: New Orleans, LA*.

Lischer, Sarah Kenyon. 2005. *Dangerous sanctuaries: refugee camps, civil war, and the dilemmas of humanitarian aid*. Cornell University Press.

Lischer, Sarah Kenyon. 2007. "Causes and Consequences of Conflict-Induced Displacement." *Civil Wars* 9(2):142–155.

Marshall, Monty G., Keith Jagers & Ted Robert Gurr. 2012. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2010." Online: <http://www.systemicpeace.org/inscr/inscr.htm>. Accessed: 11.6.2012.

Martineau, Joshua S. 2010. "Red Flags: A Model for the Early Warning of Refugee Outflows." *Journal of Immigrant & Refugee Studies* 8(2):135–157.

Melander, Erik & Magnus Öberg. 2006. "Time to Go? Duration Dependence in Forced Migration." *International Interactions* 32(2):129–152.

Melander, Erik & Magnus Öberg. 2007. "The Threat of Violence and Forced Migration: Geographical Scope Trumps Intensity of Fighting." *Civil Wars* 9(2):156–173.

Moore, Will H. & Stephen M. Shellman. 2004. "Fear of Persecution." *Journal of Conflict Resolution* 48(5):723–745.

Morrison, Andrew R. 1993. "Violence of economics: What drives internal migration in Guatemala?" *Economic Development and Cultural Change* 41:817–831.

Neumayer, Eric. 2004. "Asylum destination choice. What makes some West European countries more attractive than others?" *European Union politics* 5(2):155–180.

Newland, Kathrin. 1993. "Ethnic conflict and refugees." *Survival*.

Pini, Justin. 2008. "Political Violence and the African Refugee Experience." Online: <http://www.iar-gwu.org/node/19>. Accessed: 23.5.2012.

Rubin, Jacqueline H. & Will H. Moore. 2007. "Risk Factors for Forced Migrant Flight." *Conflict Management and Peace Science* 24(2):85–104.

Salehyan, Idean & Kristian Skrede Gleditsch. 2006. "Refugees and the Spread of Civil War." *International Organization* 60(02):335–366.

Schmeidl, Susanne. 1997. "Exploring the Causes of Forced Migration: A Pooled Time Series Analysis, 1971-1990." *Social Science Quarterly* 78(2):284–308.

Simmons, Beth A. & Zachary Elkins. 2004. "The Globalization of Liberalization: Policy Diffusion in the International Political Economy." *American Political Science Review* 98(1):171–189.

Stanley, William Deane. 1987. "Economic Migrants or Refugees from Violence? A Time-Series Analysis of Salvadoran Migration to the United States." *Latin American Research Review* 22(1):132–154.

UNHCR. 2007. "Convention and Protocol Relating to the Status of Refugees." Online: <http://www.unhcr.org/protect/PROTECTION/3b66c2aa10.pdf>. Accessed: 2.12.2010.

UNHCR. 2010. "UNHCR Statistical Online Population Database." Online: <http://apps.who.int/globalatlas/default.asp>. Accessed: 1.10.2010.

UNHCR. 2012. "UNHCR Global Trends 2011." Online: <http://www.unhcr.org/statistics>. Accessed: 1.9.2012.

UNRWA. 2010. "Number of registered refugees." Online: <http://www.unrwa.org/etemplate.php?id=253>. Accessed: 1.10.2010.

USCRI. 2009. "World Refugee Survey 2008." Online: <http://www.refugees.org/article.aspx?id=2114&subm=19&ssm=29&area=About%20Refugees>.

Vorrath, Judith & Lutz Krebs. 2009. "Democratization and Conflict in Ethnically Divided Societies." *Living Reviews in Democracy* 1.

Wahlbeck, Osten. 2002. "The concept of diaspora as an analytical tool in the study of refugee communities." *Journal of Ethnic and Migration Studies* 28(2):221–238.

Weidmann, Nils B., Doreen Kuse & Kristian Skrede Gleditsch. 2010. "The Geography of the International System: The CShapes Dataset." *International Interactions* 36(1).

Weiner, Myron. 1993. "Security, Stability, and International Migration." *International Security* 17(3):91–126.

Weiner, Myron. 1996. "Bad Neighbors, Bad Neighborhoods: An Inquiry into the Causes of Refugee Flows." *International Security* 21(1):5–42.

Wimmer, Andreas, Lars-Erik Cederman & Brian Min. 2009. "Ethnic Politics and Armed Conflict: A Configurational Analysis of a New Global Data Set." *American Sociological Review* 74(2):316–337.

Wood, William. 1994. "Forced Migration: Local Conflicts and International Dilemmas." *Annals of the Association of American Geographers* 84(4):607–634.

World Bank. 2011. "World Development Indicators 2011." Online: <http://data.worldbank.org/indicator>. Accessed: 5.06.2011.

Wucherpfennig, Julian, Nils W. Metternich, Lars-Erik Cederman & Kristian Skrede Gleditsch. 2012. "Ethnicity, the State, and the Duration of Civil Wars." *World Politics* Forthcoming.

Zhukov, Yuri M. & Brandon M Stewart. Forthcoming. "Choosing your neighbors: Networks of diffusion in international relations." *International Studies Quarterly*.

Zolberg, Aristide R., Astri Suhrke & Sergio Aguayo. 1986. "International Factors in the Formation of Refugee Movements." *International Migration Review* 20(2):151–169.

## Appendix

**Table 6:** Refugee Flows and Transnational Ethnic Linkages

<b>Refugee Flow</b>	<b>Transnational Ethnic Linkage</b>		
	No	Yes	Total
No	33'997 (92.58%)	3'991 (68.32%)	37'988 (89.25%)
Yes	2'725 (7.42%)	1'851 (31.68%)	4'576 (10.75%)
Total	36'722 (100%)	5'842 (100%)	42'564 (100%)

*Note:*  $\chi^2 = 3100$ ,  $p = 0.000$