Ethnic diversity and conflict in sub-Saharan Africa: Evidence from refugee-hosting areas*

Luisito Bertinelli† Rana Cömertpay‡ Jean-François Maystadt§

Pre-print

Abstract

This study explores how forced migration affects ethnic diversity and conflict in 23 Sub-Saharan African countries from 2005 to 2016. Using UNHCR data on refugee camp locations, we predict changes in local ethnic diversity. By integrating Afrobarometer and Ethnic Power Relations-Ethnicity of Refugees datasets, we analyze the link between refugee-induced diversity and conflict occurrence. Findings indicate that refugee-induced polarization increases the risk of local violence, while fractionalization has a mitigating effect. Notably, the number of refugees does not impact the likelihood of conflict; instead, alterations in ethnic diversity, especially polarization, emerge as the primary driver of conflict.

Keywords: Refugees, Diversity, Conflict, Migration, Africa
JEL-Classification: D74, F22, J15, O15, Q34

---

*We thank Samuel Bazzi, Michel Beine, Joseph Gomes, Arzu Kıbrıs, Anna Maria Mayda, Çağlar Özden, Lukas Delgado Prieto, Audrey Sachs, Stephen Winkler, and participants to the IFPRI Applied Microeconomics and Development Seminar, LICOS seminar at KULeuven, the Migration and Development annual conference, the Household in Conflict Network annual conference, the UNU-WIDER "The Puzzle for Peace" conference, the TREE-UPPA second workshop, Saint-Louis Economics Seminar and the Workshop in economic development at the Université Marien Ngouabi (Brazzaville). We acknowledge financial support from the World Bank Social Sustainability and Inclusion Global Practice as part of the activity "Preventing Social Conflict and Promoting Social Cohesion in Forced Displacement Contexts" under the program "Building the Evidence on Protracted Forced Displacement: A Multi-Stakeholder Partnership". The program is funded by UK aid from the United Kingdom's Foreign, Commonwealth, and Development Office (FCDO), it is managed by the World Bank Group (WBG) and was established in partnership with the United Nations High Commissioner for Refugees (UNHCR). This work does not necessarily reflect the views of FCDO, the WBG, or UNHCR.

†Associate Professor, Department of Economics and Management (DEM), University of Luxembourg, L-1359 Luxembourg. Tel: (+352) 46 66 44 6620. E-mail: luisito.bertinelli@uni.lu.

‡Research Associate, Labour Market Department, Luxembourg Institute of Socio-Economic Research (LISER), L-4366 Esch-sur-Alzette, Luxembourg. Tel: (+352) 58 58 551. E-mail: rana.coemertpay@liser.lu.

§Professor, IRES/LIDAM, UCLouvain; FNRS - Fonds de la Recherche Scientifique, Belgium; Lancaster University, Economics Department, UK. E-mail: jean-francois.maystadt@uclouvain.be.

1 Introduction

In the last decade, the number of refugees has more than doubled to over 35 million by the end of 2022 (United Nations High Commissioner for Refugees, 2023).[1] Although forced displacements in the Middle East (e.g. from Syria) and Latin America (e.g. from Venezuela) have received considerable attention, the African continent hosted 7 million refugees at the end of 2022, a number that has tripled since 2010. Many are accommodated in camps, mainly in neighbouring countries and in so-called protracted situations. While around 22% of refugees worldwide live in refugee camps (United Nations High Commissioner for Refugees, 2020), this proportion is estimated to be around 40% in sub-Saharan Africa (Verwimp and Maystadt, 2015). Organized camps remain the dominant form of accommodation in countries such as Kenya, Rwanda, Tanzania, and Uganda.[2] Over the same period, there has been a boom in the economic literature examining the impact of refugees on host economies (Ruiz and Vargas-Silva, 2013; Maystadt et al., 2019; Verme and Schuettler, 2021). The consensus is that refugees are not necessarily a burden, but can induce important distributional changes in the host population. However, the focus on labour markets may not be helpful in fully understanding the structural changes induced by these population flows. A standard narrative has been that refugees may alter the ethnic balance in their host communities, potentially leading to uncooperative behaviour and the emergence of inter-ethnic conflict (Brzoska and Frohlich, 2016; Burrows and Kinney, 2016; Mach et al., 2019, 2020). Understanding how refugees shape local diversity is crucial for assessing their developmental impact and informing resettlement policies.

This study aims to assess the impact of forced migration on ethnic diversity and conflict in host countries in sub-Saharan Africa. We combine a unique dataset on refugee camps with individual data from Afrobarometer surveys in 23 African countries for the period 2005-2016. We construct two standard measures of ethnic diversity: indices of ethnic fractionalization (EF) and ethnic polarization (EP). Ethnic fractionalization measures the probability that two individuals drawn at random from society belong to two different ethnic groups, and thus increases with the number of ethnic groups present. Ethnic polarization captures antagonism between individuals and is maximized when society is divided into two equally sized and distant ethnic groups. Although these indices have been widely used, they have shown little variation over time, making causal inference difficult.

[1] A refugee is defined as *"a person who has been forced to flee his or her country because of persecution for reasons of race, religion, nationality, political opinion or membership of a particular social group"* (1951 Convention Relating to the Status of Refugees and the 1967 Protocol, Art 1.A.2.) and includes people in refugee-like situations.

[2] We describe recent trends in forced displacement with a focus on sub-Saharan Africa in Online Appendix A.

The innovative aspect of our analysis is that we use data on the precise location of refugee camps, their annual size, and, most importantly, their annual composition in terms of countries of origin. Combined with the dataset Ethnic Power Relations - Ethnicity of Refugees 2019, we can predict annual changes in ethnic diversity induced by refugee inflows at the local level. We then assess the relationship between refugee-induced diversity and conflict intensity. Based on an intergroup competition model proposed by Esteban and Ray (2011), we argue that the presence of refugees and the resources that typically accompany such population flows are associated with a risk of intergroup conflict over these additional resources. Indeed, there is a large literature suggesting that refugees and their hosts compete over existing resources and services (Maystadt et al., 2019). In this context, we expect ethnic polarization to be much more important for conflict than ethnic fractionalization. In line with the theoretical work of Esteban and Ray (2011), we find a positive effect of ethnic polarization on the intensity (and occurrence) of conflict. A one standard deviation increase in the polarization index increases the intensity of violent conflict by 9 percentage points. This corresponds to an average increase of 9.2 percent. With less precision, the opposite effect is found for the fractionalization index. These contrasting results are confirmed by individual data on physical assault and interpersonal crime.

Our contribution is twofold. First, it has been argued that ethnic diversity is strongly linked to the lack of social cohesion in a society (Alesina et al., 2016; Arbatli et al., 2020), which can potentially result in the most extreme outcome of organized violence (Esteban and Ray, 1994, 1999; Collier and Hoeffler, 1998; Fearon and Laitin, 2003; Esteban et al., 2012b; Amodio and Chiovelli, 2018; Bazzi and Gudgeon, 2021). Other scholars have investigated intermediary outcomes, including the prevalence of mistrust (Robinson, 2017), sub-optimal provision of public goods (Habyarimana et al., 2007; Desmet et al., 2020), lower quality of institutions (Alesina et al., 1999; Alesina and Zhuravskaya, 2011), and increases in socioeconomic inequality and associated grievances. While earlier scholars considered ethnic cleavages to be deeply cultural, biological, or psychological and a deterministic driver of conflict (Horowitz, 1985; Ignatieff, 1993; Huntington, 1996), recent literature has adopted an instrumental view of the role of ethnic diversity in conflict (Ray and Esteban, 2017). We adopt a similar stance on the subject of ethnicity. Ethnic identities are multiple and malleable and can be used to facilitate strategic coordination and enforcement for collective action (Brubaker and Laitin, 1998; Esteban et al., 2012b).[3] Intuitively, a high degree of ethnic fractionalization increases the costs

[3]Ethnic identity can be defined as *"a subset of identity categories in which membership eligibility is determined by attributes associated with, or believed to be associated with descent"*. (Chandra, 2006, 298). Such descent-based attributes may encompass group characteristics such as colour, language, or religion. The implementation of this concept may prove problematic in light of the so-called aggregation problem, as outlined by (Posner, 2004). In this

of coordinating collective action and reduces the group size required for mobilization. As outlined by Esteban et al. (2012b), polarization represents a distinct phenomenon. In the context of polarized groups, the public payoff will accrue to a significant number of individuals within the winning group. It becomes more straightforward to encourage group members to engage in the conflict effort. In contrast to the extensive cross-national literature on the role of ethnic diversity (Easterly and Levine, 1997; Alesina et al., 2003; Miguel and Gugerty, 2005; Habyarimana et al., 2007), our study adopts a more recent approach, investigating similar research questions at the local level (Desmet et al., 2020; Gomes, 2020b; Montalvo and Reynal-Querol, 2021).

Furthermore, our work contributes to the recent literature on the exploitation of plausibly exogenous changes in diversity (Amodio and Chiovelli, 2018; Bazzi et al., 2019). Drawing causal inference is indeed challenging when using a time-constant measure of diversity (Montalvo and Reynal-Querol, 2005; Desmet et al., 2012), as ethnic diversity may be correlated with many unobserved characteristics. To the best of our knowledge, three other papers have recently employed time-varying changes in ethnic diversity. Firstly, Bazzi et al. (2019) investigate the impact of changes in intergroup diversity on national identity, social capital, public goods, and ethnic conflict in Indonesia. To this end, the authors examine a resettlement programme in Indonesia, elucidating the distinction between ethnic fractionalization and polarization. To that purpose, they analyze a resettlement program in Indonesia and shed light on the distinction between ethnic fractionalization and polarization. The former is associated with a greater sense of national identity (in contrast to ethnic attachment), and vice versa for the latter. Polarization is also associated with adverse effects on social capital, as evidenced by lower intergroup tolerance and trust, lower community engagement, and a proclivity for redistribution. Secondly, also relying on the case of Indonesia, Bazzi and Gudgeon (2021) use alterations in ethnic diversity, resulting from a redrawing of political boundaries, as a case study. Their findings indicate that an increase in polarization is associated with a heightened risk of violence. Thirdly, Amodio and Chiovelli (2018) investigate the impact of migration flows on ethnic diversity in South Africa following the repeal of apartheid segregation laws. They show that a greater degree of polarization among the Black population at the district level is associated with a higher incidence of armed confrontations, whereas fractionalization does not exert a discernible influence on conflict. Our paper differs from existing literature in that it specifically investigates the impact of international

paper, we adopt the approach outlined in Müller-Crepon et al. (2022) in identifying ethnic groups based on language, the most prevalent attribute globally (Gellner, 1983) and in Africa (Vail, 1989). A more detailed description of the Linking Ethnic Data from Africa (LEDA) software can be found in Section 3 and Appendix B.1.

movements of refugees on ethnic diversity in Africa. By focusing on the changes in diversity induced by non-voting migrants, we abstract from mechanisms that rely on the median voter theorem or the seizing of power (Bazzi et al., 2019; Bazzi and Gudgeon, 2021; Amodio and Chiovelli, 2018; Mayda et al., 2021). Further alternative mechanisms are discussed in greater detail in Section 6.

Second, our study contributes to another strand of the literature, which assesses the impact of forced migration on the societies in which refugees are hosted (Ruiz and Vargas-Silva, 2013; Maystadt et al., 2019; Becker and Ferrara, 2019; Verme and Schuettler, 2021). In particular, our study examines the influence of refugees in low-income countries. While the existing literature has primarily focused on the immediate effects of forced migration on labour and goods markets (Maystadt et al., 2019; Verme and Schuettler, 2021), little is known about the long-term impact on the hosting population, including the potential effects on trust and identity formation. Notable exceptions are provided by Zhou (2018), Zhou (2019), and Zhou and Shaver (2021).[4] Zhou (2018) explores the impact of refugee presence on local citizens' opposition to citizenship inclusion in sub-Saharan Africa. Zhou (2019) additionally demonstrates how the presence of refugees influences national identity formation in Tanzania, whereby local citizens may seek to differentiate themselves from a novel migrant out-group. Zhou and Shaver (2021) find that the presence of refugees does not increase the likelihood of conflict. Our study differs from these insightful studies in that we focus on the specific channel through which refugees affect these outcomes: the change in ethnic diversity. We assume that refugees are not a homogeneous group; rather, we exploit their likely ethnic attachment. While the studies by Zhou (2018, 2019) and Zhou and Shaver (2021) address different issues, they nevertheless emphasize the importance of controlling for the direct impact of refugee presence in our research design. In this regard, our findings should not be interpreted as evidence that refugees *per se* impact the probability of violence. Indeed, we do find a negative correlation between the number of refugees and the occurrence of conflict, confirming previous results from Zhou and Shaver (2021). We qualify their results by showing that refugees may increase conflict when their presence serves to intensify intergroup antagonism in communities that are already polarized.

The consideration of ethnic diversity as a mediating factor is not new. There is emerging literature suggesting that refugees and their hosts compete for available resources and services (Maystadt et al., 2019), potentially leading to security incidents along ethnic lines (Braithwaite et al., 2019). Most of

[4]The literature on social tensions between refugees and local citizens, as reviewed by Zhou et al. (2021), addresses resource competition and ethnic rivalry as contributing factors (Salehyan, 2006; Ruegger, 2017). However, the validity of this literature has been largely questioned by Zhou and Shaver (2021).

these studies are either qualitative or based on a small number of observations or they compare regions at the first administrative level in a cross-sectional analysis. For example, the role of Rwandan Hutu refugees who moved to the eastern Democratic Republic of Congo in 1994 has been largely documented in historical accounts of the so-called Congolese wars (Turner, 2007; Prunier, 2009; Stearns, 2011). The contrast with the non-conflicting role of Rwandan Hutu refugees in Tanzania can be partly explained by the polarization of ethnic identities in eastern Congo, but not in Tanzania (Whitaker, 2003). Such a contrast can also be observed between countries. While the lack of ethnic polarization is reported to explain the low likelihood of conflict in refugee-hosting countries such as Zambia and Malawi, instability is linked to ethnic rivalry in countries such as Ivory Coast, Guinea, Sierra Leone, and Uganda (Whitaker, 2003). In many contexts, ethnicity has been reported as a key conditioning factor in explaining the risk of refugee-induced political violence. Braithwaite et al. (2019, 9) argue that "refugees may contribute to the risk of political violence to the extent that they upset the ethnic balance in host states". In their empirical exercises, Fisk (2019) and Ruegger (2019) suggest that the risk of conflict in Africa is greater when refugees and the local population in host states share some ethnic ties. Our paper differs, however, by using panel data analysis, which is better suited to dealing with cross-sectional and temporal heterogeneity. We also qualify the nature of possible disturbances in ethnic diversity. Fisk (2019) finds that the presence of refugees is associated with violence, particularly when the refugees and the host region share the same ethnic group. In similar pooled estimates, co-ethnicity is approximated by the existence of common ethnic groups between source and host countries. We highlight the important distinction between ethnic fractionalization and polarization. In other words, what matters is not so much *whether* but *how* ethnic diversity changes as a result of refugee movements. Based on Esteban and Ray (2011)'s theoretical predictions, we expect ethnic polarization to be much more important for conflict than ethnic fractionalization.

The remainder of our paper is structured as follows. First, we describe our theoretical motivations in Section 2. Section 3 describes our data to provide a clear understanding of the sample used in our study. We then present our theoretical motivation and our identification strategy (Section 4). In Section 5, we first present our main results (Section 5.1). We then conduct a series of robustness tests (Section 5.2). Section 6 discusses alternative explanations, heterogeneity, and the limitations of our analysis. Section 7 concludes with policy implications.

2 Theoretical framework

We build our empirical investigation on the theoretical implications of Esteban and Ray (2011)'s game theoretic model. Esteban and Ray (2011) presents a model of competition between different groups. According to the game's unique Nash equilibrium, conflict is more likely to occur when the population is highly polarized in the presence of so-called public payoffs. On the contrary, group fractionalization is expected to play a much smaller role in this case. Ethnic fractionalization is the probability that two individuals drawn at random from society will belong to two different ethnic groups (Esteban and Ray, 1994; Esteban et al., 2012b) and is the most commonly used index to describe the ethnic structure of a society (Ray and Esteban, 2017). Assuming equal group size, the index increases with the number of groups. It captures differences in identification between groups and "reaches a maximum when everyone belongs to a different group" (Esteban et al., 2012b, 859). Given the lack of theoretical and empirical support for relating conflict and fractionalization as a measure of diversity, Esteban and Ray (1994) introduced the polarization index, which is "defined as an aggregation of all interpersonal antagonisms" (Esteban et al., 2012b, 859). This index captures the existence of deep divisions between ethnic groups and goes from zero to one but reaches its maximum when society is divided into two equal but very distant ethnic groups (Esteban and Ray, 2011). The intuition behind this theoretical prediction is that fractionalization increases the coordination costs for collective action. As described by Esteban et al. (2012b), it is different with polarization. With polarized groups, the public payoff accrues to a large number of people for the winning group, and a large number of people internalize this accrual and therefore have more incentives to contribute to the conflict effort.

In the presence of refugees and the resources that typically accompany such population flows, ethnic polarization is expected to play a more important role in intergroup conflict than ethnic fractionalization. A substantial body of literature suggests that refugees can generate new resources in local economies, particularly in the African context (Maystadt et al., 2019; Betts et al., 2017). Migrants fleeing armed conflict provide host communities with abundant and affordable labour, as seen in regions such as Darfur (Alix-Garcia et al., 2013), Rwanda (Taylor et al., 2016), Tanzania (Maystadt and Verwimp, 2014) and Uganda (Kreibaum, 2016; Kadigo and Maystadt, 2023). This influx of labour, combined with increased competition for low-skilled jobs, is leading to an increase in agricultural production and local economic growth. In Kenya, Betts et al. (2019) document the development of refugee-run markets in the Kakuma camps, the Kalobeyei settlement, and nearby towns. These thriving markets are sustained not only by refugee skills and remittances but also by economic linkages with host populations. Local markets are further expanded by refugee-host trade and increased demand from national and international humanitarian workers (Betts et al., 2019; Alix-Garcia et al., 2018). Additional resources are generated by the humanitarian sector, driven by demand for goods and services, the provision of food (some of which is exchanged on local markets), and significant investment in infrastructure such as roads, boreholes, medical facilities, and schools (Maystadt and Duranton, 2019). Theoretically, these resources can increase the potential for intergroup conflict - though not necessarily between refugees and hosts - particularly if ethnic polarization increases.[5]

3 Data

Our primary dataset combines Afrobarometer data with information on conflict occurrence and intensity and the presence and composition of refugee camps within a pre-defined 80km radius (used in the baseline results). The units of observation are 5,194 locations (Afrobarometer clusters or enumeration areas) drawn from five surveys in 23 countries (see Table D.1). The data sources used in our analysis are outlined below: Afrobarometer, UNHCR refugee camp data, Armed Conflict Location and Event Data (ACLED), Uppsala Conflict Data (UCDP), and the 2019 Ethnic Power Relations - Ethnicity of Refugees (EPR-ER) dataset. In the final subsection, we describe how these data were used to define our main variables of interest and present some descriptive statistics in Table 1.[6]

Afrobarometer. The Afrobarometer is a pan-African research network that conducts regular surveys of public attitudes on democracy, governance, the economy, and society in African countries (Afrobarometer, 2020). Using Afrobarometer's geocoded surveys, we focus on clusters as our unit of observation.[7] Our sample consists of 5,194 such places and 76,518 individuals (aged 18 and over) in 23 countries in sub-Saharan Africa. Given the sampling frame (see Appendix B.2), Afrobarometer samples are unlikely to include refugees. Afrobarometer provides geocoded data for 6 rounds,

[5]Note that the theoretical framework does not predict which group - refugees or hosts - is more likely to engage in violence. Empirically, group identification is primarily based on ethnicity.

[6]We also present descriptive statistics separately for refugee-hosting and non-refugee-hosting areas in Table D.2.

[7]Clusters correspond to location classes (administrative regions, such as states or provinces; populated places, such as towns or villages; structures, such as buildings, bridges or roads; and other topographical features, such as rivers, mountains or national parks) with exact or approximate geographic information (see Figure 1). Further details can be found in Appendix B.2.

corresponding to the period 1991-2016, with information on an individual's ethnicity available from round 3 (corresponding to 2005-2006). We therefore limit our analysis to the period 2005-2016. The selection of countries is determined by data availability (see Table D.1). Among the 32 countries with available Afrobarometer data, we exclude Botswana, Cape Verde, Lesotho, Madagascar, Mauritius, Sao Tome and Principe, South Africa, and Swaziland, for which no data on refugee camps or from the EPR-ER are available.[8] We also exclude Sudan because the question on individual ethnicity is not asked in the survey for this country. The countries in our sample are Benin, Burkina Faso, Burundi, Cameroon, Gabon, Ghana, Guinea, Ivory Coast, Kenya, Liberia, Malawi, Mali, Mozambique, Namibia, Niger, Nigeria, Senegal, Sierra Leone, Tanzania, Togo, Uganda, Zambia and Zimbabwe.

Using the Afrobarometer data has several advantages. First, similar to Berman et al. (2023), it allows us to assess local changes in ethnic diversity by matching the ethnic group information of each Afrobarometer cluster to the approximate ethnic composition of nearby refugee camps. By using the Afrobarometer data, we are using an accurate level of diversity as proxied at the time of the survey, rather than an outdated representation of ethnic composition. An alternative would be to use homeland maps based on anthropological work (e.g. Murdock). However, as Berman et al. (2023) points out, historically remote inland areas are less accurate than coastal areas.[9]

Conflict. Our paper relates variation in ethnic diversity to data on conflict from ACLED (Linke et al., 2010). Two main definitions are used: the *incidence* of conflict and the *intensity* of conflict. *Incidence* is captured by an indicator equal to one if a conflict occurred in a given year within a predefined buffer around cluster j. *Intensity* is measured by summing the number of conflict events occurring in a given year within the same buffer area. A conflict event is defined as a single altercation in which violence is used by one or more groups for a political purpose (Linke et al., 2010). We further describe events (non-exclusively) as violent events, non-violent events, violence against civilians (also called civil conflict), and riots. A refugee-related conflict event is defined as an event whose description refers to the word 'refugee'. In our main analysis, we focus on violent conflict (Section 5.1) and report results for other outcomes as robustness tests (Section 5.2). In doing so, we follow a recent and large literature that has combined the ACLED dataset with geographically disaggregated data in Africa (Besley and Reynal-Querol, 2014; Berman and Couttenier, 2015; Michaelopoulos and Papaioannou,

[8] Gambia is excluded because we only have data for the non-geocoded round 7.

[9] The Afrobarometer is also more likely to reflect changes in ethnic diversity due to endogenous indigenous displacement. We discuss this issue in Section 6.

2016; Berman et al., 2017; Harari and Ferrara, 2018; Eberle et al., 2020; McGuirk and Burke, 2020b).

As a further robustness check, we also use conflict data from the UCDP, which uses a more conservative definition of conflict. The UCDP dataset is manually curated and compiled with automated computer support (Sundberg and Melander, 2013). The UCDP defines an armed conflict event as *"an incident in which armed force was used by one organized actor against another organized actor, or against civilians, resulting in at least one direct fatality in a specific place and at a specific time"* (Pettersson et al., 2020). We extract daily event observations from the UCDP dataset if the location of the actual event is known exactly, the location of the event is within a radius of less than 25 km around a known point, or at least the administrative district where the event occurred is known. As Eberle et al. (2020) points out, the UCDP events are more likely to capture violence between large and more structured groups.

Table D.2 shows that, on average, conflict events appear to be more frequent but less intense in refugee-hosting areas. Of course, this is not a causal interpretation, but a simple correlation. As can be seen from both Panel A and Panel B, non-violent conflicts seem to occur slightly more frequently than violent ones. On average, the probability of violent conflict is about 48%, while in refugee hosting areas this rises to 52%. Conflicts between more structured and larger groups, as captured by the UCDP data, appear to be much less common.

Refugees. To exploit the variation in ethnic diversity induced by the annual variation in refugees (and also to control for the direct effect of refugees on our results), we use data on refugee camps provided by the UNHCR. The dataset contains detailed time-series information on the location and size of 1,453 refugee camps worldwide and 821 refugee camps in sub-Saharan Africa over the period 2000-2016.

To our knowledge, UNHCR currently provides the most comprehensive information on refugees at the sub-national level, which allows us to assess the ethnic composition of camps, which is key to our research question. First, in combination with the EPR-ER 2019 dataset, we use the country of origin of refugees recorded at the camp level for each year to approximate the ethnic composition of each camp. The EPR-ER records the ethnic composition of refugee stocks from neighbouring refugee-hosting countries nearby (maximum distance between country borders $\leq$ 950 km) with at least 2,000 refugees and provides the ethnic composition of refugees (Vogt and Girardin, 2015). Second, we restrict the data on refugees to those aged 18 and over to make them comparable with the Afrobarometer-based individual data. Third, we only use data on refugees hosted within the borders of the host country.

By merging the refugee camp data with the Afrobarometer, we end up with information on 172 camps located at a maximum distance of 80 km from the 5,194 clusters.[10] Figure 1 shows the locations of these refugee camps and clusters. Clusters are shown in green, while clusters near a refugee camp are shown in red. Refugee camps are marked with a red + sign.

There are some important limitations to this data. First, the data only provide information on refugees living in camps monitored by the UNHCR. As can be seen in Figure D.1, we compare the UNHCR refugee camp data on the annual number of refugees with the official UNHCR statistics on refugees (which include people in refugee-like situations) at the country level.[11] Although the overall trends are consistent, our constructed dataset underestimates the true refugee population in Africa, which is not surprising as our camp-specific data does not include dispersed refugees or refugees living outside camps.

[10] There are 189 camps at a distance of $\leq$ 120 km and 113 camps at a distance of $\leq$ 40 km from these clusters.

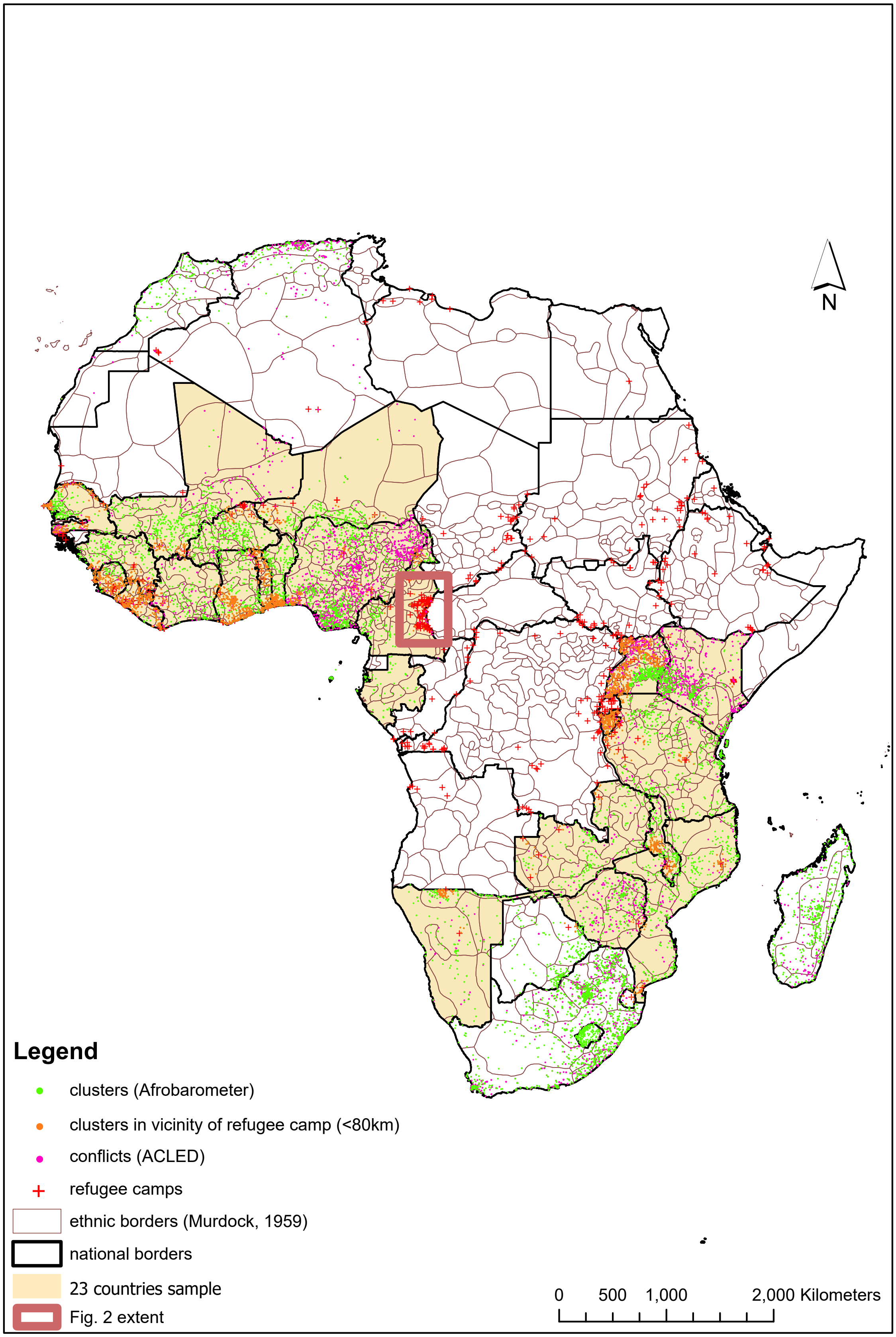
[11] The data is taken from United Nations High Commissioner for Refugees – The UN Refugee Agency (2024).

Table 1: Descriptive Statistics

VARIABLES	(1) Obs.	(2) Mean	(3) Std.Dev.	(4) Min.	(5) Max.
Conflict Events.					
Violent conflict, intensity (IHS, 80km)	14,425	0.9743	1.2662	0	6.5367
Violent conflict, incidence (80km)	14,425	0.4801	0.4996	0	1
Civilian conflict, intensity (IHS)	14,425	0.7443	1.1159	0	6.5309
Civilian conflict, incidence	14,425	0.3967	0.4892	0	1
Non-violent conflict, intensity (IHS, 80km)	14,425	1.4061	1.5442	0	5.7808
Non-violent conflict, incidence (80km)	14,425	0.5724	0.4947	0	1
Protest, intensity (IHS, 80km)	14,425	1.3006	1.5202	0	5.7746
Protest, incidence (80km)	14,425	0.5358	0.4987	0	1
UCDP conflicts, intensity (IHS)	14,425	0.2374	0.7021	0	5.5373
UCDP conflicts, incidence	14,425	0.1323	0.3389	0	1
Diversity Indices					
EF	14,425	0.1528	0.2097	0	0.8337
EP	14,425	0.0557	0.0712	0	0.25
REF (Min. Ling. Dist., 80km)	14,425	0.1651	0.2123	0	0.8337
REP (Min. Ling. Dist., 80km)	14,425	0.0596	0.0721	0	0.25
REF (no intergroup distance)	14,425	0.2820	0.2637	0	0.8664
REP (no intergroup distance)	14,425	0.1073	0.0911	0	0.25
Other variables					
Refugees (80km, IHS)	14,425	1.1159	3.1486	0	13.7611
Rain anomalies (80km)	14,425	0.2041	10.4968	-57.7804	44.8193
Temperature anomalies (80km)	14,425	0.1135	0.2204	-0.5938	1.2996

Notes: EF, EP: standard diversity indices. REF (80 km, min. ling. dist.), REP (80 km, min. ling. dist.): revised refugee diversity indices using the "minimum linguistic distance" function from LEDA. Refugees (80km, IHS): Refugees in camps in an 80-km buffer around each cluster.

Figure 1: Data and Descriptive Statistics: Clusters, Refugee Camps, and Conflicts



Combining ethnicities. A major task in constructing our dataset is to combine data on ethnicity from different sources. Indeed, linking ethnic groups is challenging because ethnic identities are socially constructed and there are different definitions, categorizations, and even conceptual approaches to identifying ethnicities in different databases or academic disciplines. This makes the task of handling, combining, and analyzing ethnicities extremely daunting, as it requires extensive background knowledge of hundreds of ethnicities, and manual handling would inevitably lead to inconsistencies, manipulation errors, or subjective decisions. Fortunately, we can rely on the open-source software package Linking Ethnic Data from Africa (LEDA), developed by Müller-Crepon et al. (2022), which contains a complete pipeline for linking ethnic datasets from Africa in a consistent and reproducible way. Müller-Crépon and his co-authors link different sources of ethnic data (including Afrobarometer, Ethnic Power Relations, and the Murdock Atlas, which we use) by using the linguistic tree from the Ethnologue and systematically grouping ethnicities. Ethnic distances are derived from this grouping in the main tree. Using the linguistic tree from the Ethnologue database, they propose a systematic solution to the problem of grouping ethnicities. More information can be found in Appendix B.1.

We obtain the ethnicity of refugees from the EPR-ER dataset, while the ethnicity of individuals in host areas comes from the Afrobarometer. These ethnicities are not systematically reported at the same level with a similar categorization process; instead, the information may refer to an individual's linguistic ethnicity, dialect, or multilingual ethnic group. In our main analysis, we use LEDA's binary linking at the "dialect" level, based on the minimum linguistic distance, to link these ethnic groups.[12] This involves computing a value corresponding to the shortest path (see equation B.1) between ethnic groups using a language tree. In our case, "dialect" is the level defined to match the two groups (see Figure D.4 from Müller-Crepon et al. (2022) for a Ghanaian case). We again rely on *LEDA* to obtain the linguistic distance between the groups, which is incorporated into the fractionalization and polarization indices described in equations (1) and (2). In line with a large literature using language as a defining attribute of ethnic identity (Desmet et al., 2012; Esteban et al., 2012b; Gomes, 2020a), (Müller-Crepon et al., 2022) uses language families, languages, and dialects from the 16th edition of the Ethnologue database (Lewis, 2009). The use of the LEDA software package is described in Appendix B.1.

[12]We also use this method to link data from the EPR-ER on the ethnicity of refugees with data from the Murdock Atlas on their historical homeland (Section 4).

Revised indices of refugee diversity. We begin by using Afrobarometer data to construct standard indices of diversity, namely the indices of ethnic fractionalization (hereafter EF) and ethnic polarization (hereafter EP) (Bazzi et al., 2019; Esteban and Ray, 1994).

The EF index describes the probability that two randomly selected individuals from a given location belong to two different ethnic groups (Alesina et al., 2003, 2016; Gomes, 2020b). The EF index can be defined as

$$EF_{jt} = \sum_{g=1}^{N_{jt}} \sum_{h=1}^{M_{jt}} s_{gt} s_{ht} d_{gh}, \tag{1}$$

where N_{jt} (M_{jt}) is the number of ethnic groups in cluster j at time t and s_{gt} (s_{ht}) is the population share of ethnic group g (h) at time t. d_{gh} is the intergroup (linguistic) distance between groups g and h. [13] Our fractionalization index incorporates intergroup distance. It is therefore similar to the Greenberg Gini index in Esteban and Ray (2011) and Esteban et al. (2012b). We include intergroup distance in both the polarization and fractionalization indices because omitting this intergroup dimension would overestimate diversity. However, our results are robust to excluding intergroup distance from our diversity indices or only from the fractionalization index (see Section 5.2). We also replicate our analysis with three separate indices similar to a specification in Esteban et al. (2012b), namely the polarization index with intergroup distance, the fractionalization index with intergroup distance (the so-called Greenberg-Gini index), and the fractionalization index without intergroup distance. Although less efficient, the results are qualitatively similar (Section 5.2).

The EP index gives more weight to between-group differences at the expense of within-group homogeneity. It can be defined as (Esteban and Ray, 1994, 1999; Montalvo and Reynal-Querol, 2005) [14]:

$$EP_{jt} = \sum_{g=1}^{N_{it}} \sum_{h=1}^{M_{jt}} s_{gt}^2 s_{ht} d_{gh}, \tag{2}$$

We compute this index for each cluster at the time of each Afrobarometer survey to assess how

[13] The Afrobarometer diversity indices use sampling weights to account for the fact that not all respondents have the same probability of being interviewed.

[14] As pointed out by Desmet et al. (2020), Esteban and Ray (1994) offer a slightly more general index, which is the formulation proposed by Reynal-Querol (2001).

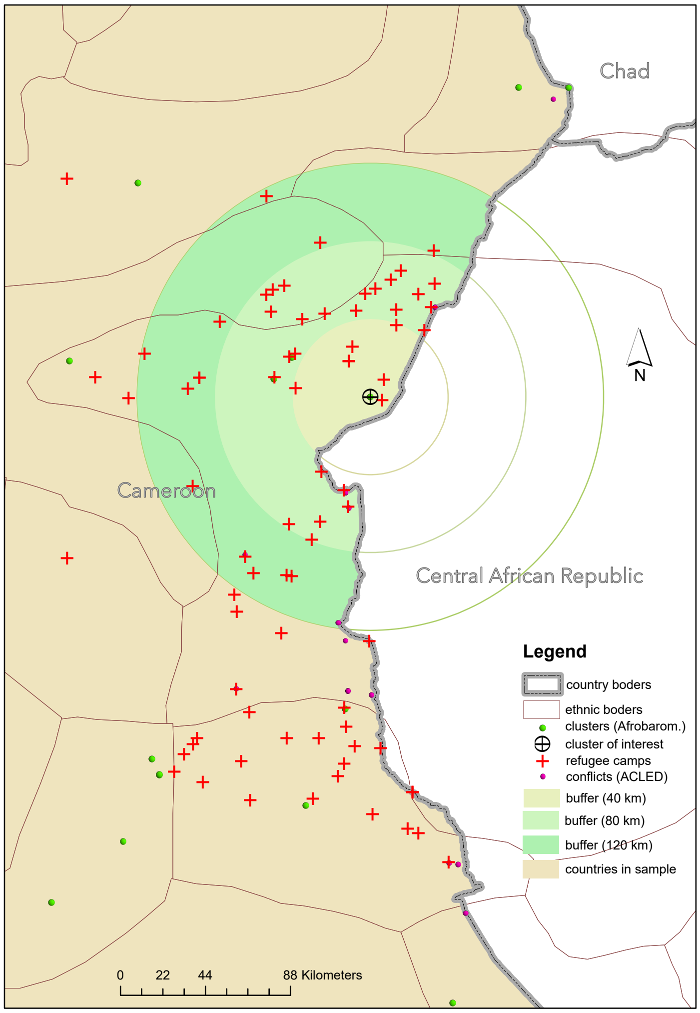
refugee-induced changes in diversity differ from standard diversity indices. To construct the revised refugee diversity indices by ethnicity g and h, we first combine information on the country of origin of refugees hosted in refugee camps c in year t with data from the EPR-ER 2019 dataset.

More specifically, the EPR-ER dataset gives us the share of refugees from ethnic group g who move from country o to country d in year t. The EPR-ER data gives us the three main ethnic groups. The number of refugees belonging to camp c in year t is therefore approximated by the following formula:

$$\sum Ref_{cgt} = Ref_{ocdt} * Share_{odgt}. \tag{3}$$

This gives us the number of refugees Ref_{cgt} of ethnicity g per camp c at time t. We can then sum the number of refugees by group g in year t for each cluster j within a buffer of, say, 80 km. To restrict the number of refugees to those within the boundaries of the host country, this buffer does not cross country borders. Figure 2 shows, for each cluster of interest (i.e. +) in our data, the refugee camps (i.e. red +) within a buffer of 40 km, 80 km, and 120 km in the host country.

Figure 2: Data and Descriptive Statistics: Computing Ethnic Diversity Indices



To obtain a sample-based number of hosts comparable to a population-based number of refugees, we multiply the number of respondents in the Afrobarometer surveys by the ratio $\frac{\mathcal{N}}{n}$, where $\mathcal{N}$ is the total population of the surveyed country in year t and n is the sample size of the survey in year t. Intuitively, we need to ensure that a refugee is comparable to a host in the calculation of the revised diversity indices. Based on the World Development Indicators, a country's population aged 18 and over in $t-1$ is considered equivalent to the most recent official national census used as the sampling frame for Afrobarometer surveys. Since individuals in the Afrobarometer are aged 18 and over, we also restrict our analysis to refugees aged 18 and over for comparative purposes.[15]

There is a limitation to our approximation in equation 3. The ethnic composition of refugees

[15]On average, 45% of refugees in camps are aged 18 and over.

in each year t for a given country of origin-destination pair obtained from the EPR-ER database is assumed to be homogeneous across camps of the same country of origin-destination pair for refugees in year t. This may seem a strong assumption; however, the risk of misallocating refugees is reduced because the annual variation in the EPR-ER is generated by only a few dominant groups for a given origin-destination pair, and the geographical distribution of refugees by country of origin is strongly influenced by proximity to their countries of origin.[16]

As can be seen from Table 1, in refugee-hosting areas, on average, both EF and EP seem to increase quite significantly when they are *revised* by including the number of refugees in an 80 km buffer: the mean value of the standard EF index is 15.28%, while the mean value of the revised refugee EF index is 16.51%. The mean of the standard EP index is 5.57%, while the mean of the revised refugee EP index is 5.96%. While Figure 3 shows considerable variation in both indices within our sample when averaging these indices at the regional level over the study period, Figure 4 also shows substantially higher variability in the refugee-corrected diversity indices compared to the original indices (with the 45-degree line measuring the equality between refugee-corrected and uncorrected indices).

[16]It is possible that our approximation is noisy and could potentially introduce non-random measurement errors. In Section 4 we propose an instrumental variable approach and estimate $Share_{odgt}$ from the EPR-ER data using a gravity model. Our findings on the number of ethnic groups over time for a given origin-destination pair are consistent with the EPR-ER data. It appears that refugees of a given origin-destination pair mainly belong to two major ethnic groups. This also means that the variation in diversity in refugee hosting areas is due to the composition of refugees at the camp level. Figure D.2 shows the movements of refugees from the countries of origin to the countries of destination under consideration. Somalia, the Democratic Republic of Congo, Liberia, South Sudan, and Sudan are the main countries of origin of refugees, while Kenya, Tanzania, Uganda, Zambia, and Ghana appear to be the countries hosting the most refugees. Figure D.3, which shows refugees in camps by ethnicity for the top 5 countries of asylum over the sample period, shows that there is considerable variation in the ethnic composition of camps.

Figure 3: Ethnic Fractionalization and Ethnic Polarization

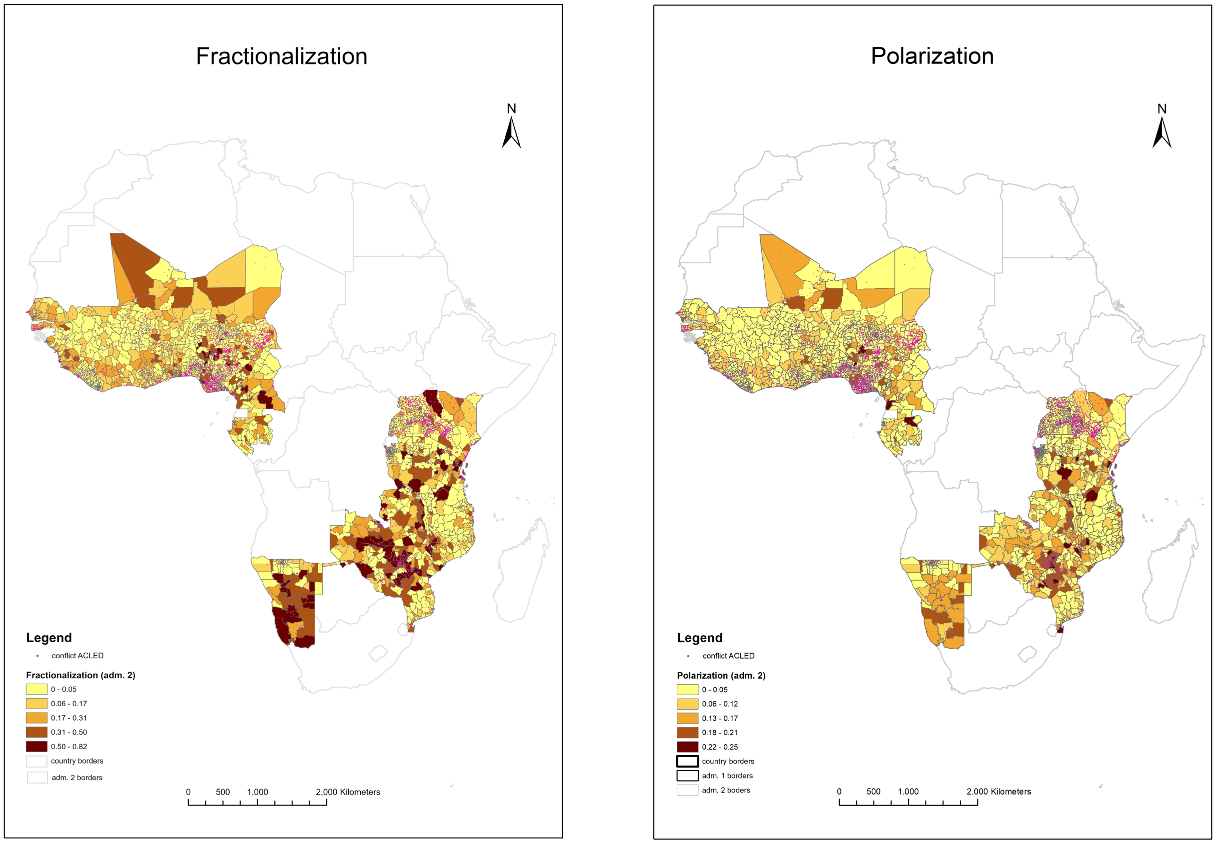
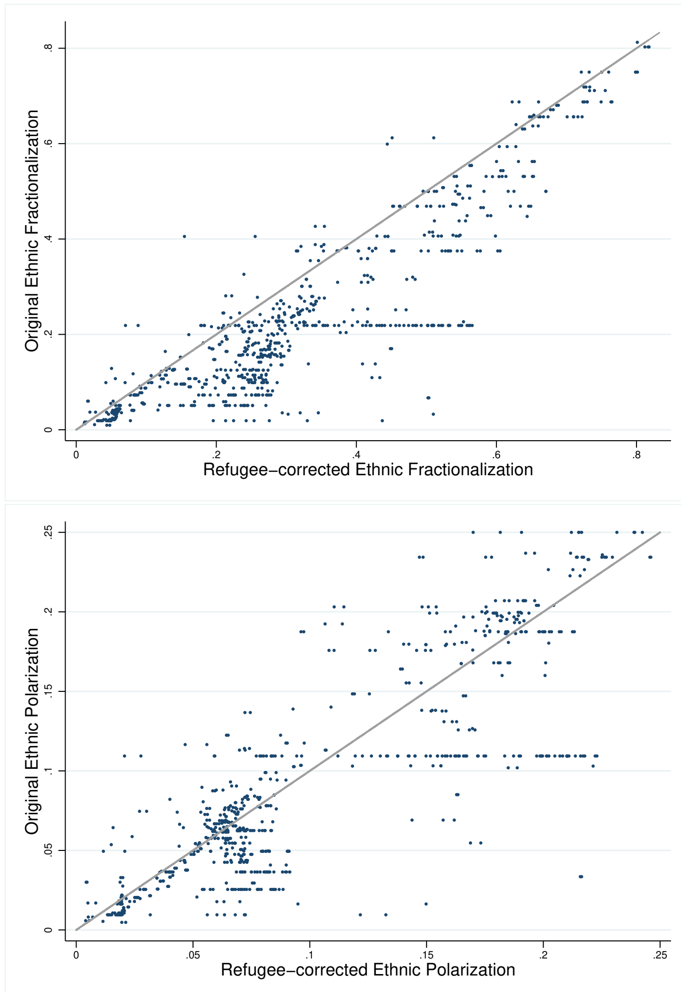


Figure 4: Original vs Refugee-corrected measures of Ethnic Fractionalization and Polarization



Note: Refugee corrected diversity indices are based on 80km buffers; all clusters with at least two different ethnicities and in the vicinity of a refugee camp are included.-

4 Identification strategy

Main specifications. In line with the theoretical predictions described in Section 2, we aim to assess how refugee-induced ethnic fractionalization and ethnic polarization affect conflict in the following way:

$$Conflict_{jt} = \alpha_j + \tau_t + \beta_1 REF_{jt-1} + \beta_2 REP_{jt-1} + \beta_3 Refugees_{jt-1} + \beta_4 Q_{jt} + \epsilon_{jt}, \quad (4)$$

where $Conflict_{jt}$ represents the incidence or intensity of conflict in location j in year t. We define locations using information from the Afrobarometer Enumeration Areas. In the remainder of this paper, we refer to the Afrobarometer Enumeration Areas as clusters.

The number of conflict events is transformed into the inverse hyperbolic sine to facilitate interpretation (Bellemare and Wichman, 2020). REF_{jt} and REP_{jt} refer to *revised refugee ethnic fractionalization* and *revised refugee ethnic polarization* respectively. They capture the change in ethnic diversity induced by refugee inflows at time $t-1$ and are constructed by revising standard diversity measures based on the Afrobarometer with the changes in ethnic diversity induced by refugee flows within a specified buffer around each cluster j. This is the main departure from the existing literature, as we revise standard measures of ethnic diversity to include the annual variation in refugee ethnicity. We then construct a measure of proximity between clusters in the host country and refugees in surrounding camps by defining an 80 km buffer around each cluster.[17] This is quite close to what is commonly used in the existing literature on conflict risk, which constructs grids of 50 x 50 km and their immediately adjacent cells, essentially creating concentric rings with radii of 50, 100, and 150 km (Tapsoba, 2022). Furthermore, this choice of buffer size ensures that between 75 percent and virtually all refugee camps fall within a cluster buffer. Other studies using Afrobarometer data construct buffers ranging from 25 km (e.g. (Michaelopoulos and Papaioannou, 2011), examining ethnic-specific pre-colonial institutional structures) to 100 km (e.g. (McGuirk and Burke, 2020a), analyzing the impact of food price shocks on conflict).

To control for unobserved heterogeneity and changes within a given cluster, we introduce cluster and year fixed effects, α_j and δ_t. To minimize the risk of confounding the refugee-induced changes in diversity with the annual changes in refugee numbers, we also control for the presence of refugees based on the same buffer as the one used to construct the refugee-induced change in diversity. More specifically, the variable $Refugees_{jt-1}$ counts the number of refugees present in cluster j in year $t-1$ within the predefined buffer. The variable is also transformed into an inverse hyperbolic sine to ease interpretation.

Finally, Q_{jt} controls for yearly shocks at the cluster level, such as weather shocks. In particular, we control for rain and temperature anomalies.[18] Standard errors are clustered at the Afrobarometer

[17] We test the robustness of our results with a smaller (40 km) and a larger (120 km) radius in Section 5.2

[18] We employed data sourced from the Climatic Research Unit (CRU) at the University of East Anglia, specifically using the CRU TS Version 4.04 dataset (Harris et al., 2020). The dataset comprises rainfall and temperature information, which were obtained at a spatial resolution of 0.5 x 0.5 degrees in longitude $\times$ latitude. To facilitate our analyses, we calculated the mean values of the climatic variables within designated buffers. Subsequently, anomalies were computed to discern deviations from the established climatic norms (subtracting the long-term mean and dividing

cluster level.

Identification threats. Our main concern relates to the endogeneity of our refugee-induced diversity indices. There are two sources of endogeneity in assessing the impact of refugees. First, the location of refugee camps is not random. Most of them are located in peripheral areas and their location could be correlated with unobserved characteristics related to the conflict (Ruiz and Vargas-Silva, 2013; Maystadt et al., 2019; Becker and Ferrara, 2019; Verme and Schuettler, 2021). Therefore, we do not ascribe a causal interpretation to the *direct* presence of refugees. We only check the sensitivity of our results to the addition or omission of such a "bad control". The second source of endogeneity relates to the non-random distribution of refugees. We cannot rule out the possibility that refugees are non-randomly sorted into areas with particular ethnic characteristics.[19] Qualitatively, we should first acknowledge that the ability of refugees to choose their place of residence is very limited (Maystadt and Verwimp, 2014) and that "campsite selection operates based on several rather time-invariant factors" (Salemi, 2021, 6). Factors such as proximity to the border or availability of land should be captured by the site fixed effects. Comparing our diversity indices, correcting or not for the annual variation in refugees, suggests that refugees increase both indices (see Table 1). If minimizing security risk and maximizing cultural proximity between refugees and hosts are possible objectives of campsite selection - as argued by Salemi (2021) - we would expect our results for refugee-induced ethnic fractionalization and polarization indices to be biased downwards.

To assess this potential endogeneity threat, we also provide an instrumental variable (IV) approach as a robustness check. In particular, we are concerned that certain ethnic groups from certain countries of origin move to destination countries with similar ethnic characteristics. Such endogenous selection would be reflected in the EPR-ER data. To construct a plausibly exogenous instrumental variable, we first implement a gravity model to predict the number of refugees of a given ethnic group e moving from country o to d at time t, based on the EPR-ER data. The intuition behind this zero-step gravity model is to predict the number of refugees from a particular ethnic group moving to a particular destination based on the proximity of the historical homeland of that ethnic group in the country of origin and past conflicts in the historical homeland of that ethnic group and the countries of origin. The plausibly exogenous number of refugees from a given ethnic group moving from one

by the standard deviation).

[19] Another source of sorting may arise from refugees gravitating towards countries with more liberal admission policies (Blair et al., 2022), but it is not clear how this would affect ethnic diversity.

country to another is then used as an input, similar to equations 1, 2 and 3, to construct predicted refugee-induced diversity indices, which are used as instrumental variables in a two-stage framework. Further details can be found in the Online Appendix C.

5 Results

In this section, we discuss the results of our benchmark analysis (Section 5.1), and of several robustness tests with alternative outcome variables and alternative specifications (Section 5.2).

5.1 Main Results

Table 2 presents the results of a linear model with the intensity of violent conflict as the dependent variable. Following the theoretical framework proposed by Esteban and Ray (2011), we include both diversity indices in the same specification. Columns (1) and (2) present the standard diversity indices, while Columns (3) to (6) present our revised refugee diversity indices. In Columns (2), (4) and (6) we take into account the presence of refugees within a radius of 80 km. Although recent literature has downplayed the conflicting effects of refugees in host areas (Zhou and Shaver, 2021), the magnitude of our coefficients may be affected by the confounding effects of refugee presence. In addition, Columns (5) and (6) include climate controls, with Column (6) representing our benchmark specification based on equation 4.

Columns (1) and (2) show that without accounting for changes in ethnic diversity induced by refugees, we would not be able to identify a relationship between diversity and violent conflict. In Column (3), the revised index of refugee fractionalization has a negative and significant coefficient, while the revised index of refugee polarization has a positive and significant effect on the intensity of violent conflicts. In Columns (4) and (6), the coefficient of the revised refugee polarization index is of the same order of magnitude when the number of refugees is controlled for. The coefficient is only estimated at the 10 percent level - possibly due to the local nature of our approach to diversity (see Section 5.2) - but is economically significant (see Section 6). Fractionalization does not matter, provided there are refugees. Refugees *per se* do not directly exacerbate conflict intensity. On the contrary, we find a negative correlation between the presence of refugees and the dependent variable. Although correlational, this is not surprising given the existing literature on compensating economic effects (Maystadt et al., 2019; Verme and Schuettler, 2021; Alix-Garcia et al., 2018) that materialize, for example, through labour markets (Buscher and Vlassenroot, 2009; Maystadt and Verwimp, 2014; Ruiz and Vargas-Silva, 2016, 2015), improved consumption (Maystadt and Verwimp, 2014; Kreibaum, 2016; Taylor et al., 2016; Alloush et al., 2017; Foltz and Shibuya, 2022), improved infrastructure (Maystadt and Duranton, 2019) or the provision of local public goods (Maystadt and Verwimp, 2014; Kreibaum, 2016; Zhou et al., 2021). The importance of polarization is not altered by the inclusion of rainfall and temperature anomalies (Columns (5) and (6)). According to our benchmark specification in Column (6), an increase of one standard deviation (0.0721) in the revised EP index of refugees increases the intensity of violent conflict by 9 percentage points. At the mean, this corresponds to an increase of about 9.2 percent. Without contradicting Zhou et al. (2021), our results offer an alternative explanation: it is not the size of refugee flows that matters, but the way they change the ethnic composition of refugee-hosting areas. More specifically, the risk of conflict increases when refugees increase the ethnic polarization of host communities.

Table 2: Benchmark Analysis: Diversity and Violent Conflict

	(1)	(2)	(3)	(4)	(5)	(6)
			Violent Conflict, Intensity			
Native EF	-0.1677	-0.1306				
	(0.2759)	(0.2780)				
Native EP	1.1930	1.0849				
	(0.7468)	(0.7524)				
Refugees (80km, IHS)		-0.0175***		-0.0184***		-0.0183***
		(0.0059)		(0.0062)		(0.0062)
REF (Min. Ling. Dist., 80km)			-0.4357*	-0.2920	-0.4534*	-0.3114
			(0.2600)	(0.2666)	(0.2621)	(0.2686)
REP (Min. Ling. Dist., 80km)			1.3600**	1.2229*	1.3864**	1.2526*
			(0.6824)	(0.6865)	(0.6882)	(0.6919)
Rain anomalies (80km)					0.0017**	0.0016**
					(0.0008)	(0.0008)
Temp anomalies (80km)					-0.0919**	-0.0971**
					(0.0450)	(0.0450)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.807	0.807	0.807	0.807	0.807	0.807
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS, reported in Column (6). Columns (1) and (2) present the standard diversity indices. From Column (3) onwards, the revised refugee diversity indices are presented. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80 km buffer around each cluster is used to *revise* the standard ethnic diversity measures with the number of refugees in camps within that distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between the Afrobarometer and EPR-ER datasets. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.
* More information on LEDA in Appendix B.1.

5.2 Robustness

We now examine the sensitivity of these main results to a series of robustness checks.[20] First, we discuss the assumption of homogeneity of treatment effects. Second, we assess the robustness of our results to the use of alternative outcomes (Table 3). Third, we conduct robustness tests with some alternative specifications (Table 4).

Alternative transformations. Chen and Roth (2023), Mullahy and Norton (2022) and Aihounton and Henningsen (2020) show that the regression results are not independent of the units of measurement of the IHS-transformed variables. In other words, it is not independent of scaling. Alternative transformations for unit-independent estimates have been proposed, such as conversion to a rank with some reference distribution (Delius and Sterck, 2024; Chen and Roth, 2023) or the use of power transformations (Thakral and To, 2023). Table D.4 shows results with alternative transformations. To compare effect sizes, we calculate the quasi-elasticity, expressed as the estimated coefficient multiplied by the associated standard deviation and divided by the mean of the dependent variable. We confirm a positive and significant coefficient for the refugee-induced polarization index and a negative coefficient for the refugee-induced fractionalization index. The latter is only marginally significant for higher power transformations. Our estimated quasi-elasticities for the refugee-induced polarization index range from 0.04 to 0.17 across specifications. Since Aihounton and Henningsen (2020) recommend the use of R-squared and predicted R-squared as criteria for model selection, we keep the IHS transformation for our main results and conclude with a quasi-elasticity of around 0.09 for the refugee-induced polarization index.

Alternative Outcomes. Each row in Table 3 corresponds to the same specification as in Table 2, with one alternative outcome. Due to space constraints, only the results for the revised refugee EF and EP -our variables of interest- from Column (6), corresponding to Equation 4, are presented.[21] Row A presents the results of our benchmark estimation in Column (6) of Table 2. Transformed into

[20] Recent research has highlighted the shortcomings of the two-way fixed effects estimator (TWFE) in the presence of heterogeneous treatment effects (Athey and Imbens, 2021; Borusyak et al., 2021; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021). No new methods have yet been developed to deal with a continuous treatment variable in unbalanced panel data. Nevertheless, we can downplay the risk of heterogeneous treatment effects in our setting. We adopt the main diagnostic test proposed by Jakiela (2021). Defining the treated group as areas receiving refugees, we regress the residual outcome (partialled out with place and year fixed effects) on the residual treatment variables (partialled out with place and year fixed effects). We find no evidence that the treatment effects are heterogeneous across clusters and years (Table D.3).

[21] Notes below Table 3 include cross-references to tables in the Appendix with the full set of specifications corresponding to each row.

an inverse hyperbolic sine, our results can then be interpreted as quasi-elasticities. As can be seen from Table 3, the revised refugee EP has a significant impact on the other conflict measures, except non-violent conflict (Rows C and D), civilian conflict incidence (Row F), protest intensity (Row G) and refugee-related conflict incidence (Row J). The magnitude of the coefficients for civil conflict intensity is particularly high (Row E). The results for the revised refugee EP are also confirmed when using the incidence of violent conflict (Row B) or protests (Row H). The results indicate that a one standard deviation increase in polarization is associated with a 3.6 percentage point increase in the probability of violent conflict. Interestingly, our results do not hold for violence perpetrated by larger and more structured groups, as captured by the UCDP data (lines K and L).

Alternative diversity indices. As can be seen from Table 4, our results do not depend on the choices made in the construction of our main variables of interest. Here, each row represents a modification of our benchmark specification, corresponding to equation 4, and is run in the same way as in Table 2. Again, for reasons of space, only the results for the revised refugee EF and EP - our variables of interest - from Column (6) are presented. Row A presents the results from our benchmark estimation in Column (6) of Table 2. We find similar results when all groups are assumed to be equally distant, i.e. the indices do not include the linguistic distance between groups (Row B), or when the linguistic distance between groups is only omitted in the fractionalization index (Row C).[22] We also replicate our analysis with three separate indices similar to a specification in Esteban et al. (2012b), namely the polarization index with intergroup distance, the fractionalization index with intergroup distance (the so-called Greenberg-Gini index), and the fractionalization index without intergroup distance. Although less efficient, the results are qualitatively similar (Row D).

Alternative aggregations. First, the size of the buffer used to capture the number of refugees around clusters and their contribution to ethnic composition is rather arbitrary, so we tested other buffer sizes. Our results are robust to using both a smaller (40 km) and a larger (120 km) buffer, as shown in Rows E and F.[23]

[22] In this case, as described by Esteban et al. (2012a,b), the fractionalization index can be denoted as $\sum_{g=1}^{N_{jt}} \sum_{h \neq 1}^{M_{jt}} s_{gt} s_{ht} = \sum_{g=1}^{N_{jt}} s_{gt}(1 - s_{gt})$. The polarization index, without intergroup distance, can be denoted as $\sum_{g=1}^{N_{jt}} \sum_{h \neq 1}^{M_{jt}} s_{gt}^2 s_{ht}$.

[23] The size of the buffer is largely an arbitrary choice. For example, Salemi (2021) uses a buffer of 30 kilometres when studying the impact of refugee camps on deforestation in Africa, while Maystadt et al. (2020) investigates the same research question with coarser cells of 110 kilometres over 110 kilometres. While there is a strong argument for using fine spatial resolution when studying an outcome such as deforestation - which is likely to be inversely related to camp proximity - it is less obvious for our study to focus on conflict dynamics. In a related context of violence prediction,

Second, we should recognize that diversity indices are more likely to be measured with noise in highly diverse communities at the local level.[24] Despite the use of sampling weights in the construction of the diversity indices to take into account the fact that not all respondents had the same probability of being interviewed, we have no guarantee that our diversity indices are representative at the local level. Although similar ethnic diversity indices have been used at the local level (Nunn and Wantchekon, 2011; Rohner et al., 2013; Robinson, 2017; Desmet et al., 2020; Gomes, 2020b,a; Hodler et al., 2020), we cannot exclude the possibility that a lack of representativeness at the local level introduces some noise into our estimates. Ideally, we would have liked to construct our local diversity indices based on census data. However, such data are not available on an annual basis, and only a minority of African countries include questions on ethnicity in their censuses (Robinson, 2017). We argue, however, that such concerns should not be overstated. Indeed, such noise cannot easily explain the contrast between the coefficients corresponding to the pre-revised and revised indices and the opposite results found for the revised refugee fractionalization and the revised polarization. This set of results can be explained by the fact that our identification is based on annual changes in refugee flows. In addition, we also aggregate the number of conflict events at the regional level at the cost of introducing an attenuation bias.[25] Row G confirms the negative and positive effects found for the revised fractionalization and polarization indices.[26] Third, we do not only include clusters from the Afrobarometer with precision code 2, as in the rest of the analysis but keep all clusters for which we have geographical references. The qualitative results are the same, but we lose statistical significance (Row H).

Tapsoba (2022) relies on the equivalent of 50km grid cells to account for conflict events, recognizing that conventional methods in the literature consider such a resolution, which is analogous to delineating concentric rings at intervals of 50, 100 and 150 kilometres. Negative and positive coefficients for refugee-induced fractionalization and polarization are found when a smaller buffer is used, but the corresponding coefficients are far from being precisely estimated. One reason for this is that the proportion of treated observations drops from about 12 percent with an 80 km buffer (16% with a 120 km buffer) to less than 6% with a 40 km buffer, less than 2% with a 20 km buffer, and less than 1% with a 10 km buffer.

[24] Robinson (2017) highlights another risk that ethnic diversity may also capture different theoretical mechanisms at aggregated levels.

[25] Note that the IV approach is likely to deal with the measurement errors if they are correlated with our main variables of interest. Our IV estimates therefore capture a local average treatment effect arising from the plausibly exogenous increase in annual refugee flows of particular ethnic groups. The similarity of the IV results to the OLS results supports this interpretation.

[26] The regional aggregation implemented in Table D.21 uses the GADM2 classification corresponding to the second sub-national administrative division. The name of this type of administrative unit may vary from one country to another but can be considered as corresponding to the district level. It has an average size of 4,651 km^2, with a minimum of 0.636 km^2 (Southern Law in Nigeria) and a maximum of 345,145 km^2 (Tombouctou in Mali).

Alternative models. As in our benchmark analysis, we have so far used a linear model. We also perform robustness tests using a non-linear conditional logit model (Row I). This non-linear estimation does not change the essence of our results. Despite the use of location fixed effects, another concern might be that areas with historically high levels of polarization might react differently to other shocks in terms of conflict risk. In Row J we show the most complete specification, where we control for historical polarization (using the Murdock data) interacted with year fixed effects. Our results are robust to this specification.

Ethnic salience. Ethnic identity may become more salient through the experience of conflict. Since similar ethnic groups are likely to share common borders (Michaelopoulos and Papaioannou, 2016), it is not impossible that conflict could spill over through this channel. To control for the risk of conflict spillovers, Row K of Table 4 shows that our results are robust to the addition of an interaction term between the distance to the border and the year fixed effects.

Endogenous location of refugees The literature recognizing the endogenous location of refugees is ubiquitous (Ruiz and Vargas-Silva, 2013; Maystadt et al., 2019; Becker and Ferrara, 2019; Verme and Schuettler, 2021). In our case, refugee camps are mainly located in peripheral areas, which are likely to have different characteristics. With the use of location fixed effects, this is less of a concern in our research design, unless such selection is correlated with the ethnic composition of refugee hosting areas. This would be the case if refugees were self-selected into areas with co-ethnic residents. Despite the plausibility of our identification assumptions, we cannot rule out the possibility that refugees sort ethnically. As an additional analysis, we therefore implement the 2SLS approach, using the results of a gravity equation to predict where refugees would go based only on plausibly exogenous factors. Row L of Table 4 confirms the positive effect found for the revised refugee polarization index. Although the negative sign is preserved, no significant coefficient is found for the equivalent fractionalization index.[27]

[27] We report our detailed results in Table D.26.

Table 3: Summary Table: Alternative Outcomes

	(1)	(2)
	REF	REP
A. Benchmark results (N=14,425)[a]	-0.3114	1.2526*
	(0.2686)	(0.6919)
B. Violent conflict, incidence (N=14,425)[b]	-0.1958*	0.4965*
	(0.1145)	(0.2953)
C. Non-violent conflict, intensity (N=14,425)[c]	-0.1785	0.5807
	(0.2757)	(0.7164)
D. Non-violent conflict, incidence (N=14,425)[d]	-0.0931	0.4711
	(0.1105)	(0.3002)
E. Civilian conflict, intensity (N=14,425)[e]	-0.3279	1.5113**
	(0.2432)	(0.6019)
F. Civilian conflicts, incidence (N=14,425)[f]	-0.0882	0.4592
	(0.1105)	(0.2849)
G. Protest, intensity (N=14,425)[g]	-0.1205	0.5514
	(0.2702)	(0.7006)
H. Protest, incidence (N=14,425)[h]	-0.1306	0.5879*
	(0.1118)	(0.3024)
I. Refugee-related conflict, intensity (N=14,425)[i]	-0.0260	0.3062**
	(0.0481)	(0.1292)
J. Refugee-related conflict, incidence (N=14,425)[j]	0.0080	0.2363
	(0.1027)	(0.2754)
K. Conflict (UCDP), intensity (N=14,425)[k]	-0.1219	0.1559
	(0.2108)	(0.5089)
L. Conflict (UCDP), incidence (N=14,425)[l]	0.0255	-0.0928
	(0.0769)	(0.1903)

Notes: Estimated equation: Equation (4) using OLS with alternative dependent variables. Level of analysis, countries, period, and LEDA function: similar to Table 2. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. Robust standard errors clustered at the cluster level are reported in brackets. FE: fixed effects.

See Column (6) in the following tables for: [a] in Table 2; [b] in Table D.5; [c] in Table D.6; [d] in Table D.7; [e] in Table D.8; [f] in Table D.9; [g] in Table D.10; [h] in Table D.11; [i] in Table D.12; [j] in Table D.13; [k] in Table D.14 and [l] in Table D.15.

Table 4: Summary Table: Alternative Specifications

	(1)	(2)
	REF	REP
A. Benchmark results (N=14,425)[a]	-0.3114	1.2526*
	(0.2686)	(0.6919)
B. No intergroup dist. (N=14,425)[b]	-0.3105	0.9221*
	(0.2087)	(0.5186)
C. No intergroup dist. for EF (N=14,425)[c]	-0.1735	0.9029**
	(0.1270)	(0.4184)
D. Controlling for EF without intergroup dist. (N=14,425)[d]	-0.1579	1.1741*
	(0.3127)	(0.6911)
E. Buffer at 40 km (N=14,425)[e]	-0.2999	1.0642*
	(0.2188)	(0.5823)
F. Buffer at 120 km (N=14,425)[f]	-0.4867*	1.3839*
	(0.2801)	(0.7127)
G.Aggregation at the GADM2 level (N=1,563)[g]	-1.5886**	4.0304*
	(0.7993)	(2.1914)
H. Incl. all geocoded locations (N=23,236)[h]	-0.1327	0.7087
	(0.1878)	(0.4728)
I. Non-linear model (N=5,749)[i]	-0.2478*	0.6157*
	(0.1441)	(0.3610)
J. Historical Ethnic polarization (N=14,425)[j]	-0.2874	2.5072**
	(0.2700)	(0.9912)
K. Conflict Spillovers (N=14,425)[k]	-0.3036	1.2360*
	(0.2728)	(0.7062)
L. Instrumenting refugee location (N=14,425)[l]	-0.2815	1.3493*
	(0.2816)	(0.7357)

Notes: Estimated equation: Equation (4) by OLS with alternative specifications, except for Row I, estimated by LOGIT. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. Robust standard errors clustered at the cluster level are shown in parentheses.

See Column (6) for: [a] in Table 2; [b] in Table D.16; [c] in Table D.17; [d] in Table D.18; [e] in Table D.19; [f] in Table D.20; [g] in Table D.21; [h] in Table D.22; [i] in Table D.23; [j] in Table D.24; [k] in Table D.25 and [l] in Table D.26.

6 Discussion

In this section, we discuss the economic significance of our findings and draw some policy implications by discussing some alternative explanations and socio-demographic aspects available from the Afrobarometer data.

Economic significance. We estimate that, on average, the intensity and frequency of violent conflict increase by 9 percentage points (9.2 percent at the mean) and 3.6 percentage points, respectively, following a refugee-induced increase in polarization by one standard deviation. Our results confirm the positive effect of polarization on conflict found in the existing literature, although not to the same extent. Amodio and Chiovelli (2018) found that a one standard deviation increase in polarization was associated with more than four times the average number of conflicts before the democratic transition in South Africa. Their closest estimates (col. 7 of their Table 3) would give an increase of 204 percentage points (10.15*0.201*100). Such a large effect is certainly due to the particular context and regime change experienced in South Africa in the early 1990s. The magnitude of our results is more in line with other studies. Our effect size on the extensive margin of conflict compares well with Bazzi and Gudgeon (2021). These authors find that a one standard deviation increase in polarization implies a 3.6 percentage point higher probability of social conflict, about half of the cross-country correlation between ethnic polarization and low-intensity civil conflict in Esteban et al. (2012a). Our magnitude is thus almost identical to Bazzi and Gudgeon (2021), but lower than the corresponding increases of 4.15 percentage points and 13.4 percentage points found in Bazzi et al. (2019) and Amodio and Chiovelli (2018).[28]

Although not significantly different from zero in the most complete specifications, we find a negative effect on the coefficient of the revised refugee fractionalization index. When examining the extensive conflict margin, a marginally significant negative coefficient is found, corresponding to a decrease of 4.2 percentage points following a one standard deviation increase in fractionalization (0.212). The negative sign found for the fractionalization index contrasts with Esteban et al. (2012a) and Esteban et al. (2012b). Our result remains consistent with studies using local variation in diversity. Bazzi et al. (2019) estimate a similar decrease of 4.2 percentage points following an increase in fractionalization by one standard deviation, while Amodio and Chiovelli (2018) report a decrease of 6.8 percentage points. Bazzi and Gudgeon (2021) also finds a negative coefficient, but it is far

[28]Based on Column 4 of Table 11 in Bazzi et al. (2019) and Column 5 of Table 3 in Amodio and Chiovelli (2018).

from being statistically different from zero. As pointed out by Bazzi et al. (2019), the negative sign is still consistent with Esteban and Ray (2011) in a situation where conflict materializes over public resources. Indeed, Esteban et al. (2012a) and Esteban et al. (2012b) interpret the positive coefficient found for fractionalization as evidence that private components, such as the existence of natural resources, matter. In our context, where refugees are likely to be accompanied by contested public resources, intergroup contact with many small groups (high fractionalization) is likely to reduce the risk of conflict.

Finally, the magnitude of our results can also be compared with other important determinants of conflict in Africa, namely the role of economic shocks (often associated with climatic shocks), natural resources, and price shocks.[29] First, one of the most robust findings is the link between economic shocks and conflict in Africa (Blattman and Miguel, 2010). The seminal paper by Miguel et al. (2004) shows that a 1 standard deviation increase in economic growth reduces the likelihood of conflict by more than 16 percentage points.[30] Although the income effect is not the only possible interpretation (Mach et al., 2019), economic shocks in Africa have often been associated with climatic shocks (Harari and Ferrara, 2018), and the meta-analysis by Hsiang et al. (2013) suggests that a 1 standard deviation change in climate towards warmer temperatures or more extreme rainfall increases the incidence of conflict by 14 percent on average. A second important determinant of violence in Africa is the presence of natural resources. For example, Berman et al. (2017) found that a 1 standard deviation increase in the price of minerals is associated with a 5.6 percentage point increase in conflict. But other price shocks also matter. For example, a 1 standard deviation increase in producer prices reduces conflict intensity by 17 percent at the median (McGuirk and Burke, 2020b). The equivalent change in consumer prices increases conflict by 8 percent. According to Berman and Couttenier (2015), a one standard deviation increase in world demand for agricultural commodities also increases conflict by 1 to 3 percentage points. Overall, while we do not claim to provide an exhaustive review of the literature, our estimated effect sizes are smaller but comparable to other major determinants of conflict in Africa.

[29] Since Collier and Hoeffler (1998) and Fearon and Laitin (2003) there has been a booming literature on the economics of conflict. We restrict our comparison to time-varying factors. Indeed, it is difficult to compare with long-term (time-constant) determinants of conflict such as ethnic partition (Michaelopoulos and Papaioannou, 2016) or historical conflict (Besley and Reynal-Querol, 2014).

[30] We should acknowledge that the validity of these results beyond 1999 has been questioned (Ciccone, 2011; Miguel and Satyanath, 2011).

Alternative explanations. The theoretical framework used to guide this analysis is mainly driven by competition for resources between (ethnic) groups. Social conflict is mainly driven by a combination of intergroup differences and intragroup cohesion. Such a theory implies that polarization is more likely to capture intergroup antagonism and competition between a few large groups. In this way, our results suggest that refugee influxes affect group sizes, diversity indices, incentives to compete for resources, and thus conflict. Alternative explanations may challenge the fixed nature of groups and the distances between them. For example, Bazzi et al. (2019) shows that polarization increases ethnic attachment. Others have emphasized the reduction of trust, either interpersonal or institutional (Alesina and Ferrara, 2002; Beugelsdijk and Klasing, 2016).

To assess the importance of alternative explanations, we first replicate our analysis using individual data on violence. In addition to participation in protests, we follow McGuirk and Burke (2020b) in using Afrobarometer survey data on interpersonal crime and physical assault. We then assess the relationship between the revised refugee diversity indices and alternative individual outcomes such as ethnic vs. national identity, generalized trust, trust in neighbours, and institutional trust (trust in government).[31] To do so, we adopt a similar specification to our benchmark estimation at the individual level:

$$Y_{ijt} = \alpha_j + \tau_t + \gamma_1 REF_{jt-1} + \gamma_2 REP_{jt-1} + \gamma_3 Refuges_{jt-1} + \gamma_4 X_{ijt} + \gamma_5 Q_{jt} + \nu_{ijt}, \tag{5}$$

where Y_{ijt} represents several outcomes such as the likelihood of experiencing assault, crime, or theft, participating in a protest, ethnic (vs. national) attachment, interpersonal, neighbourhood, and institutional trust of individual i in cluster j surveyed in year t. The other variables are similar to equation 4, except for X_{ijt}. X_{ijt} is a vector of individual control variables such as age, education, sex, marital status, and rural/urban status. To assess the risk of inappropriate controls, we introduce these control variables progressively. Sampling weights are used to make our estimates representative at the country level.[32]

As shown in Table D.29, the results on physical assault confirm our main findings, but with a much larger magnitude.[33] We find that a one standard deviation increase in the revised ethnic polarization index increases the probability of experiencing physical assault by 5.9 percentage points. Such a change is equivalent to an increase of about 70 percent at the mean. Although similar in magnitude,

[31] Questions used as proxies for these outcomes are described in Appendix B.2.
[32] We provide descriptive statistics for these variables in Table D.28.
[33] Detailed results are provided in Table D.30.

the estimated coefficient for the revised ethnic polarization index is not statistically different from zero for interpersonal crime. Table D.29 also shows that none of our coefficients of interest statistically affect the other individual outcomes of ethnic attachment, generalized trust, trust in neighbours, and trust in institutions.

Heterogeneity. One way of identifying possible entry points for policy intervention is to exploit the socio-demographic heterogeneity provided by individual data. Indeed, the conflict literature suggests that the likelihood of engaging in violence is negatively correlated with age, female gender, wealth, and employment. This is generally attributed to higher opportunity costs for the elderly, women, and wealthier segments of the population (Blattman and Miguel, 2010). To shed light on particular vulnerabilities, we assess the impact of the revised indices of ethnic fractionalization and polarization on samples grouped according to these individual characteristics. Our results are shown in Figure D.5. Our analysis certainly lacks the power to detect clear heterogeneous effects. However, it appears that our revised indices of refugee polarization and fractionalization are stronger when the respondent is unemployed and aged between 18 and 25. Other heterogeneous effects are much less clear-cut. We remain cautious about the lack of power in this analysis, but the unemployed and the young may be a particular target for interventions aimed at reducing prejudice and increasing cooperation between groups.

Overall, our results challenge previous findings on the impact of refugees on conflict, suggesting that the relationship depends largely on how diversity changes as a result of refugee influx. When polarization increases, the risk of conflict increases. The opposite is true for fractionalization. It is therefore important to consider that the risk of conflict increases when refugees tend to increase polarization between a few large groups. In this case, fostering intergroup interaction will not necessarily reduce intergroup prejudice and increase cooperation as in other contexts (Finseraas and Kotsadam, 2017; Corno et al., 2022). Given the lack of results for alternative outcomes such as trust or ethnic attachment, competition between polarized groups is the most likely driver of our results. Implementing specific interventions in refugee-hosting and polarized communities is therefore strongly recommended.[34]

[34] See Paluck (2012) for a review of possible interventions to reduce prejudice and conflict. For example, intergroup sports have been shown to help rebuild intergroup social cohesion among displaced people in northern Iraq (Moussa, 2020).

Limits. It should be noted that many other factors may influence conflict, and we cannot definitively exclude the possibility that they are not correlated with our main variables of interest. Regarding the most established determinants of conflict, such as the role of economic shocks (often associated with climatic variations) and the presence of natural resources, we doubt that these would move simultaneously with the variation in diversity induced by refugee flows. Although the influx of new resources associated with the influx of refugees may exacerbate the public nature of the prize to be fought over, beyond its effect through the diversity indices, any increased incentive will be captured by the variable related to the size of the refugee population. We are more concerned about the role of economic inequality, migration, or price shocks. Economic inequality could interact with ethnic diversity in two ways.[35] On the one hand, economic inequality between groups is likely to exacerbate distance between groups, which explains why "horizontal inequalities" are strongly correlated with conflict (Stewart, 2000; Ostby, 2008; Cederman et al., 2011). On the other hand, inequality within groups has been shown to affect the likelihood of violence. Theoretically, groups with higher levels of in-group inequality are more effective in conflict (Esteban and Ray, 2011; Esteban et al., 2012a). The lack of socio-economic information on refugees prevents us from testing these theoretical predictions.[36] Second, anecdotal evidence suggests that refugee camps have often attracted internal migrants from surrounding areas (Maystadt and Verwimp, 2014; Maystadt and Duranton, 2019). Although there is little empirical support for migration as a driver of conflict (Fearon and Laitin, 2011; Mach et al., 2019, 2020), a link between in-migration and conflict could potentially influence our results. While the direct effect of refugees is accounted for by controlling for the number of refugees in our empirical specification, internal migrants can certainly affect the ethnic diversity of the refugee-hosting area. Such an indirect effect is likely to be captured by our previously revised fractionalization and polarization indices, but we cannot distinguish the role of internal migrants in

[35] Heavily influenced by prominent writers such as Karl Marx or Montesquieu, class struggle - or, more generally, economic inequality - has been argued to be a major driver of conflict (Gurr and Harff, 1991). Even Sen (1997, 1) suggests that "the relationship between inequality and rebellion is indeed a close one, and it runs both ways. That a perceived sense of inequality is a common ingredient of rebellion in societies is clear enough." However, this relationship lacks empirical support (Russett, 1964; Muller, 1985; Midlarski, 1988; Lichbach, 1989). More recently, (Collier and Hoeffler, 2004) and (Fearon and Laitin, 2003) found that income inequality does not systematically affect the risk of conflict. Empirically, this may be explained by the fact that economic inequality does not change sufficiently over time to explain variation in violence. Theoretically, the lack of relationship between income inequality and conflict can be explained by the fact that the have-nots do not have the means to organize violence (Esteban et al., 2012a; Ray and Esteban, 2017). As summarized by (Ray and Esteban, 2017, 276), "the rich have the means but not the motive to express this conflict, while the poor have the motive but lack the means".

[36] We can use wealth information from the Afrobarometer to approximate location-based inequality using the method proposed by McKenzie (2005). We do not find a stronger association between polarization and conflict in unequal areas. However, these results are not presented (available on request) as they do not capture in-group inequalities in the absence of socio-economic information on refugees.

the effects of the revised refugee diversity indices. Finally, refugee flows have been found to have an inflationary effect in host regions (Alix-Garcia and Saah, 2010; Maystadt et al., 2019). While this is certainly an important policy dimension, there is no reason to believe that such price effects are correlated with our diversity indices.

In addition, the external validity of our findings is limited by data availability. We only use UNHCR-monitored camps due to data availability. This means that our results may not apply to other forms of forced displacement, such as internally displaced persons (IDPs) or so-called dispersed refugees in informal settlements or urban areas. It also means that our findings are less relevant for countries that allow refugees to move freely.[37] Finally, both our theoretical framework and our data analysis cannot identify refugees or hosts as perpetrators of violence. Group identification can only be based on ethnicity. More fundamentally for policy, we cannot quantify the nature of interactions between refugees and their hosts and their mediating role in changing ethnic diversity and conflict.

7 Conclusions

Refugees have often been blamed for fuelling social conflict in their host countries. Previous research has rejected a causal effect of refugee hosting on violence (Zhou and Shaver, 2021). We provide further insight by highlighting a particular channel through which refugees may affect levels of violence in their host communities, namely the resulting changes in ethnic composition. We use annual variations in the presence of refugees to approximate the resulting changes in the diversity of refugee-hosting areas in 23 countries in sub-Saharan Africa between 2005 and 2016. Our results point to the risk of conflict when refugees exacerbate ethnic polarization in host communities. A one standard deviation increase in the polarization index increases the intensity of violent conflict by 9 percentage points (3.6 percentage points for incidence), equivalent to a change of about 9 percent at the mean. Our estimated effect sizes are comparable in magnitude to other determinants of conflict, such as economic, price, and climatic shocks. It is therefore important for policymakers and practitioners to consider that the risk of violence increases when refugees exacerbate polarization between a few large groups. In this case, promoting intergroup interaction will not necessarily reduce intergroup prejudice and increase cooperation, as has been found in other contexts (Finseraas and Kotsadam, 2017; Corno

[37]For example, Uganda has had a unique refugee policy since 2009. Broadly speaking, the Uganda Refugee Act gives refugees the "right" to move, work, and do business.(Betts et al., 2017, 2019; Kadigo and Maystadt, 2023). As shown in Table D.31, our results are nevertheless robust to the omission of this particular country.

et al., 2022). The identification of specific interventions in polarized refugee-hosting communities is therefore urgently needed.

Others, such as Betts et al. (2023), have identified programmes that seek to improve social cohesion between refugees and hosts, such as facilitating opportunities for refugee-host interactions, promoting intragroup attitudinal change, focusing on perceived "winners" and "losers" among hosts, or designing social cohesion programmes in urban and camp settings. There is also a burgeoning literature attempting to assess the role of cash transfers, education programmes, and targeted assistance in mitigating social tensions between displaced and host populations, but so far with mixed results (Aguero and Fasola, 2021; Ferguson et al., 2022; Lehmann and Masterson, 2020). Our paper shows that these efforts should be targeted primarily at highly polarized host areas. A mapping exercise combining indices of fractionalization and polarization at the local level with information on the ethnic composition of new refugee flows could provide valuable information for policymakers and organizations seeking to implement initiatives to improve refugee-host relations. Such an exercise would also require the systematic collection of ethnic information in future UNHCR and World Bank surveys.

The results obtained using individual data also highlight the importance of ethnic diversity in refugee reception situations. Of particular interest is the fact that our indices of polarization and fractionalization are stronger in explaining the likelihood of protesting when the respondent is unemployed and young. This population may therefore be a particular target for interventions aimed at reducing prejudice and increasing intergroup cooperation. For example, we know that cash transfer programmes have been particularly effective (compared to skills training and microfinance, for example) in promoting employment and social stability in poor and fragile states (Blattman and Ralston, 2015). The increased cooperation between UNHCR and other development and peace-building actors since the Global Compact on Refugees provides an appropriate framework for supporting such interventions.

References

Afrobarometer (2020). Afrobarometer Data, [Algeria, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Egypt, Gabon, Gambia, Ghana, Guinea, Ivory Coast, Kenya, Lesotho, Liberia, Morocco, Madagascar, Malawi, Mali, Mauritius, Mozambique, Namibia, Niger, Nigeria, São Tomé and Príncipe, Senegal, Sierra Leone, South Africa, Swaziland, Tanzania, Togo, Uganda, Zambia and Zimbabwe, 1-6, 1991-2016, available at http://www.afrobarometer.org.

Aguero, J. and E. Fasola (2021). Distributional Policies and Social Cohesion in a High-Unemployment Setting. Unpublished Working paper. Commissioned as part of the "Preventing Social Conflict and Promoting Social Cohesion in Forced Displacement Contexts" Series, Washington, DC: World Bank Group.

Aihounton, G. B. D. and A. Henningsen (2020, 10). Units of measurement and the inverse hyperbolic sine transformation. The Econometrics Journal 24(2), 334–351.

Alesina, A., R. Bakir, and W. Easterly (1999). Public goods and ethnic divisions. The Quarterly journal of Economics 114(4), 1243–1284.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. journal of Economic Growth 8(2), 155–194.

Alesina, A. and E. L. Ferrara (2002). Who trusts others? Journal of Public Economics 85(2), 207–234.

Alesina, A., S. Michaelopoulos, and E. Papaioannou (2016). Ethnic Inequality. Journal of Political Economy 124, 428–488.

Alesina, A. and E. Zhuravskaya (2011). Segregation and the Quality of Government in a Cross-Section of Countries. American Economic Review 101, 1872–1911.

Alix-Garcia, J., A. Bartlett, and D. Saah (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 D. Saah (2010). The Effect of Refugee Inflows on Host Communities: Evidence from Tanzania. World Bank Economic Review 24(1), 148–170.

Alix-Garcia, J., S. Walker, A. Bartlett, H. Onder, and A. Sanghi (2018). Do refugee camps help or hurt hosts? The case of Kakuma, Kenya. Journal of Development Economics 130, 66–83.

Alloush, M., J. Taylor, A. Gupta, R. Rojas Valdes, and E. Gonzales-Estrada (2017). Economic life in refugee camps. World Development 95, 334–347.

Amodio, F. and G. Chiovelli (2018). Ethnicity and violence during democratic transitions: Evidence from South Africa. journal of the European Economic Association 16(4), 1234–1280.

Arbatli, C. E., Q. H. Ahsraf, O. Galor, and M. Klemp (2020). Diversity and conflict. Econometrica 88(2), 727–797.

Athey, S. and G. Imbens (2021). Design-based Analysis in Difference-in-Differences Settings with Staggered Adoption. Journal of Econometrics.

Bazzi, S., A. Gaduh, A. Rothernberg, and M. Wong (2019). Unity in Diversity? How Intergroup Contact can Foster Nation Building. American Economic Review 109(11), 3978–4025.

Bazzi, S. and M. Gudgeon (2021). The Political Boundaries of Ethnic Division. American Economic journal: Applied Economics.

Becker, S. O. and A. Ferrara (2019). Consequences of forced migration: A survey of recent findings. Labour Economics 59, 1–16.

Beine, M., S. Bertoli, and J. F.-H. Moraga (2016). A practitioners' guide to gravity models of international migration. The World Economy 39(4), 496–512.

Bellemare, M. and C. Wichman (2020). Elasticities and the inverse Hyperbolic Sine Transformation. Oxford Bulletin of Economics and Statistics 82(1), 0305–9049.

BenYishay, A., R. Rotberg, J. Wells, Z. Lv, S. Goodman, L. Kovacevic, and D. Runfola (2017). Geocoding Afrobarometer Rounds 1-6: Methodology & Data Quality. AidData.

Berman, N. and M. Couttenier (2015). External shocks, internal shots: The geography of civil conflicts. Review of Economics and Statistics 97(4), 758–776.

Berman, N., M. Couttenier, and V. Girard (2023). Mineral resources and the salience of ethnic identities. Economic Journal 133, 1705–1736.

Berman, N., M. Couttenier, D. Rohner, and M. Thoenig (2017). The Mine is Mine! How minerals fuel conflicts in Africa. Amercian Economic Review 107(6).

Besley, T. and M. Reynal-Querol (2014). The legacy of historical conflict: Evidence from Africa. American Political Science Review 108(2), 319–336.

Betts, A., L. Bloom, J. Kaplan, and N. Omata (2017). Refugee economies: Forced displacement and development. Oxford University Press.

Betts, A., A. Chaara, N. Omata, and O. Sterck (2019). Refugee economies in Uganda: What difference does the self-reliance model make. Oxford Refugees Studies Centre.

Betts, A., A. Delius, C. Rodgers, O. Sterck, and M. Stierna (2019). Doing Business in Kakuma: Refugees, Entrepreneurship, and the Food Market. Oxford Refugees Studies Centre.

Betts, A., M. F. Stierna, N. Omata, and O. Sterck (2023). Refugees welcome? Inter-group interaction and host community attitude formation. World Development 161.

Beugelsdijk, S. and M. Klasing (2016). Diversity and trust: The role of shared values. Journal of Comparative Economics 44, 522–540.

Blair, C. W., G. Grossman, and J. M. Weinstein (2022). Liberal displacement policies attract forced migrants in the global south. American Political Science Review 116(1), 351–358.

Blattman, C. and E. Miguel (2010). Civil War. Journal of Economic Literature 48(1), 3–57.

Blattman, C. and L. Ralston (2015). Generating Employment in Poor and Fragile States: Evidence from Labor Market and Entrepreneurship Programs. Mimeo.

Borusyak, K., X. Jaravel, and J. Spiess (2021). Revisiting event study designs: Robust and efficient estimation. Mimeo.

Braithwaite, A., I. Salehyan, and B. Savun (2019). Refugees, forced migration, and conflict: Introduction to a special issue. Journal of Peace Research 56(1), 5–11.

Brubaker, R. and D. Laitin (1998). Ethnic and nationalist violence. Annual Review of Sociology 24, 423–452.

Brzoska, M. and C. Frohlich (2016). Climate change, migration and violent conflict: vulnerabilities, pathways and adaptation strategies. Migration and Development 5(2), 190–210.

Burrows, K. and P. Kinney (2016). Exploring the climate change, migration and conflict nexus. Environmental Research and Public Health 13(443), 1–17.

Buscher, K. and K. Vlassenroot (2009). Humanitarian Presence and Urban Development: New Opportunities and Contrasts in Goma, DRC. Disasters 34(2), 256–273.

Cederman, L., N. Weidmann, and K. Gledisch (2011). Horizontal inequalities and ethno-nationalist civil war: a global comparison. American political Science Review 105, 478–495.

Chandra, K. (2006). What is ethnic identity and does it matter? Annual Review of Political Science 9(1), 397–424.

Chen, J. and J. Roth (2023, 12). Logs with Zeros? Some Problems and Solutions. The Quarterly Journal of Economics, qjad054.

Ciccone, A. (2011). Economic shocks and civil conflict: A comment. American Economic Journal: Applied Economics 3(4), 215–227.

Collier, P. and A. Hoeffler (1998). On Economic Causes of Civil War. Oxford Economic Papers 50, 563–573.

Collier, P. and A. Hoeffler (2004). Greed and Grievance in Civil War. Oxford Economic Papers 56, 563–595.

Corno, L., E. L. Ferrara, and J. Burns (2022). Interaction, Stereotypes and Performance: Evidence from South Africa. American Economic Review 112(12), 3848–75.

Crozet, M. (2004). Do Migrants Follow Market Potentials? An Estimation of a New Economic Geography Model. Journal of Economic Geography 4(4), 439–458.

de Chaisemartin, C. and X. D'Haultfœuille (2020, September). Two-way fixed effects estimators with heterogeneous treatment effects. American Economic Review 110(9), 2964–96.

Delius, A. and O. Sterck (2024). Cash transfers and micro-enterprise performance: Theory and quasi-experimental evidence from kenya. Journal of Development Economics 167, 103232.

Desmet, K., J. Gomes, and I. Ortuno (2020). The geography of linguistic diversity and the provision of public goods. Journal of Development Economics 143, 102384.

Desmet, K., I. Ortuno-Ortin, and R. Wacziarg (2012). The political economy of linguistic cleavages. Journal of Development Economics 97, 322–338.

Devictor, X., Q.-T. Do, and A. Levchenko (2021). The globalization of refugee flows. Journal of Development Economics 150, 102605.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: Policies and ethnic divisionsn. Quarterly Journal of Economics 112(4), 1203–1250.

Eberle, U., D. Rohner, and M. Thoenig (2020). Heat and Hate: Climate security and farmer-herd conflicts in Africa . CEPR Discussion paper 15542.

Esteban, J.-M., L. Mayoral, and D. Ray (2012a). Ethnicity and conflict: An empirical study. American Economic Review 102(4), 1310–1342.

Esteban, J.-M., L. Mayoral, and D. Ray (2012b). Ethnicity and conflict: Theory and facts. Science 336, 858.

Esteban, J.-M. and D. Ray (1994). On the Measurement of Polarization. Econometrica 62(4), 819–851.

Esteban, J.-M. and D. Ray (1999). Conflict and Distribution. Journal of Economic Theory 87, 379–415.

Esteban, J.-M. and D. Ray (2011). Linking conflict to inequality and polarization. American Economic Review 101(4), 1345–1374.

Fearon, J. D. and D. D. Laitin (2003). Ethnicity, Insurgency, and Civil War. American Political Science Review 97(1), 75–90.

Fearon, J. D. and D. D. Laitin (2011). Sons of the Soil, Migrants, and Civil War. World Development 39(2), 199–211.

Ferguson, N., R. W. anf Laila Amine, E. Ramadi, and L. Shahin (2022). Building Stability Between Host and Refugee Communities: Evidence from a TVET Program in Jordan and Lebanon. World Bank Policy Research Working Paper 10101.

Finseraas, H. and A. Kotsadam (2017). Does personal contact with ethnic minorities affect anti-immigrant sentiments? Evidence from a field experiment. European Journal of Political Research 56(3), 703–722.

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

Foltz, J. and S. Shibuya (2022). The Effects of Internally Displaced Peoples on Income and Inequality in Mali. World Bank Policy Research Working Paper 10054.

Garcia, A. J., D. K. Pindoliay, K. K. Lopiano, and A. J. Tatemzz (2015). Modeling internal migration flows in sub-Saharan Africa using census microdata. Migration Studies 3(1), 89–110.

Gellner, E. (1983). Nations and Nationalism. Ithaca, NY: Cornell University Press.

Gomes, J. (2020a). Linguistic fractionalization and health information in Sub-Saharan Africa. World Bank Economic Review 34(1), S20–S25.

Gomes, J. (2020b). The health costs of ethnic distance: evidence from sub-Saharan Africa. Journal of Economic Growth 25, 195–226.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics 225(2), 254–277.

Gurr, T. R. and B. Harff (1991). Ethnic Conflict in World Politics, Dilemmas in World Politics. Oxford University Press, Oxford.

Habyarimana, J., M. Humphreys, D. N. Posner, and J. M. Weinstein (2007). Why does ethnic diversity undermine public goods provision? American Political Science Review 101(4), 709–725.

Harari, M. and E. L. Ferrara (2018). Conflict, climate, and cells: A disaggregated anlysis. Review of Economics and Statistics 100(4), 594–608.

Harris, I., T. Osborn, P. Jones, et al. (2020). Version 4 of the cru ts monthly high-resolution gridded multivariate climate dataset. Scientific Data 7(1), 109.

Hodler, R., S. Srisuma, A. Vesperoni, and N. Zurlinden (2020). Measuring ethnic stratification and its effect on trust in Africa. Journal of Development Economics 146, 102475.

Horowitz, D. (1985). Ethnic Groups in Conflict. Berkeley: University California Press.

Hsiang, S., M. Burke, and E. Miguel (2013). Quantifying the influence of climate on human conflict. Science 342(6151), 1235367.

Huntington, S. (1996). The Clash of Civilizations and the Remaking of World Order. New York: Simon and Schuster.

Ignatieff, M. (1993). Blood and Belonging. London: Noonday Press.

Jakiela, P. (2021). Simple dagnostcs for two-way fixed effects. Mimeo.

Kadigo, M. M. and J.-F. Maystadt (2023). How to cope with a refugee population? evidence from uganda. World Development 169, 106293.

Kreibaum, M. (2016). Their suffering, our burden? How Congolese refugees affect the Ugandan population. World Development 78, 262–87.

Lehmann, C. and D. Masterson (2020). Does Aid Reduce Anti-Refugee Violence? Evidence from Syrian Refugees in Lebanon. American Political Science Review 114(4), 1335–1342.

Lewis, P. (2009). Ethnologue: Languages of the world volume 16. SIL International Dallas, TX.

Lichbach, M. I. (1989). An Evaluation of "Does economic inequality breed political conflict?" Studies. World Politics 41(4), 431–470.

Linke, Raleigh, C. A., H. Hegre, and J. Karlsen (2010). Introducing ACLED-Armed Conflict Location and Event Data. Journal of Peace Research 47(5), 651–660.

Mach, K., N. Adger, H. Buhaug, M. Burke, J. Fearon, C. Field, C. Hendrix, C. Kraan, J. Maystadt, J. O'Loughlin, P. Rossler, J. Scheffran, K. Schultz, and N. von Uexkull (2020). Directions for reserach on climate and conflict. Earth's Future 8, 1–7.

Mach, K., C. Kraan, N. Adger, H. Buhaug, M. Burke, J. Fearon, C. Field, C. Hendrix, J. Maystadt, J. O'Loughlin, P. Rossler, J. Scheffran, K. Schultz, and N. von Uexkull (2019). Climate as a risk factor for armed conflict. Nature 571, 193–197.

Mayda, A. M. (2010). International migration: A panel data analysis of the determinants of bilateralflows. Journal of Population Economics 23(4), 1249–1274.

Mayda, A. M., G. Peri, and W. Steingress (2021). The political impact of immigration: Evidence from the us. American Economic Journal: Applied Economics 11(14), 1–33.

Maystadt, J.-F. and G. Duranton (2019). The Development Push of Refugees. Journal of Economic Geography 19(2), 299–334.

Maystadt, J.-F., K. Hirvonen, A. Mabiso, and J. Vandercasteelen (2019). Impacts of Hosting Forced Migrants in Poor Countries. Annual Review of Resource Economics 11, 439 – 459.

Maystadt, J.-F., V. Mueller, J. Van Den Hoek, and S. van Weezel (2020). Vegetation changes attributable to refugees in africa coincide with agricultural deforestation. Environmental Research Letters 15(4), 044008.

Maystadt, J.-F. and P. Verwimp (2014). Winners and Losers Among a Refugee-Hosting Population. Economic Development and Cultural Change 62(4), 769–809.

McGuirk, E. and M. Burke (2020a). The economic origins of conflict in africa. Journal of Political Economy 128(10), 3940–3997.

McGuirk, E. and M. Burke (2020b). The economic origins of conflict in Africa. Journal of Political Economy 128(10), 3940–3997.

McKenzie, D. (2005). Measuring Inequality with Asset Indicators. Journal of Population Economics 18(2), 299–360.

Michaelopoulos, S. and E. Papaioannou (2011). Divide and rule or the rule of the divided? Evidence from Africa. NBER Working Paper Series 17184, 1–77.

Michaelopoulos, S. and E. Papaioannou (2016). The long-term effects of the scamble for Africa. American Economic Review 106(7), 1802–1848.

Midlarski, M. (1988). Rulers and the ruled: patterned inequality and the onset of mass political violence. American Political Science Review 82(2), 491–509.

Miguel, E. and M. Gugerty (2005). Ethnic diversity, social sanctions, and public goods in Kenya. Journal of Public Economics 89(11), 2325–2368.

Miguel, E. and S. Satyanath (2011). Re-examining economic shocks and civil conflict. American Economic Journal: Applied Economics 3(4), 228–232.

Miguel, E., S. Satyanath, and E. Sergenti (2004). Economic Shocks and Civil Conflict : An Instrumental Variable Approach. Journal of Political Economy 112(4), 725.

Montalvo, J. G. and M. Reynal-Querol (2005). Ethnic polarization, potential conflict and civil war. American Economic Review 95(3), 796–816.

Montalvo, J. G. and M. Reynal-Querol (2021). Ethnic diversity and growth: Revisiting the evidence. Review of Economics and Statistics 103(3), 521–532.

Moussa, S. (2020). Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. Science 369, 866–870.

Mullahy, J. and E. Norton (2022). Why transform Y? A critical assessment of dependent-variable transformations in regression models for skewed and sometimes-zero outcomes. NBER working paper series 30735.

Muller, E. N. (1985). Income Inequality, Regime Represiveness, and Political Violence. American Sociological Review 50(1), 47–61.

Murdock, G. (1967). The ethnographic atlas: a summary. Ethnology 6, 109–236.

Müller-Crepon, C., Y. Pengl, and N.-C. Bormann (2022). Linking ethnic data from africa (leda). Journal of Peace Research 59(3), 425–435.

Nunn, N. and L. Wantchekon (2011). The Slave Trade and the Origins of Mistrust in Africa. American Economic Review 101(7), 3221–3252.

Ostby, G. (2008). Polarization, horizontal inequalities and violent civil conflict. Journal of Peace Research 45, 143–162.

Paluck, E. (2012). Interventions aimed at the reduction of prejudice and conflict. The Oxford Handbook of Intergroup Conflict Linda Tropp, ed., (Oxford: Oxford Library of Psychology), 400.

Pettersson, T., S. Högbladh, and M. Öberg (2020). Organized Violence. Journal of Peace Research. 1989-2019 57(4).

Posner, D. (2004). Measuring ethnic fractionalization in Africa. American journal of Political Sciences 48(4), 849–863.

Prunier, G. (2009). Africa's World Wars: Congo, the Rwandan Genocide, and the Making of a Continental Catastrophe. Oxford: Oxford University Press.

Ravenstein, E. (1885). The Laws of Migration. Journal of the Statistical Society of London 48(2), 167–235.

Ravenstein, E. G. (1989). The laws of migration – second paper. Journal of the Royal Statistical Society 52(2), 241–305.

Ray, D. and J.-M. Esteban (2017). Conflict and Development. Annual Review of Economics 9, 263–293.

Reynal-Querol, M. (2001). Religious and ethnic conflict, political systems and growth. Ph. D. thesis, London School of Economics.

Robinson, A. L. (2017). Ethnic Diversity, Segregation and Ethnocentric Trust in Africa. British Journal Political Sciences 50, 217–239.

Rohner, D., M. Thoenig, and F. Zilibotti (2013). Seeds of trust: Conflict in Uganda. Journal of Economic Growth 18, 217–252.

Ruegger, S. (2017). Refugees and Conflict Diffusion. Peace and Conflict, 142–153.

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

Ruiz, I. and C. Vargas-Silva (2013). The Economics of Forced Migration. Journal of Development Studies 49(6), 772–784.

Ruiz, I. and C. Vargas-Silva (2015). The Labor Market Impacts of Forced Migration. American Economic Review: Papers & Proceedings 105(5), 581–586.

Ruiz, I. and C. Vargas-Silva (2016). The labour market consequences of hosting refugees. Journal of Economic Geography 16(3).

Russett, B. M. (1964). Inequality and Instability: The Relation of Land Tenure to Politics. World Politics 16(3), 442–454.

Salehyan, I. (2006). Refugees and the Spread of Civil War. International Organization 60(2), 335–366.

Salemi, C. (2021). Refugee camps and deforestation in sub-saharan africa. Journal of Development Economics 152, 102682.

Sen, A. (1997). On Economic Inequality. Oxford: Clarendon Press.

Stearns, J. (2011). Dancing in the Glory of Monsters: The Collapse of the Congo and the Great War of Africa. Public Affairs.

Stewart, F. (2000). Crisis Prevention: Tackling Horizontal Inequalities. Oxford Development Studies 28(3), 245–262.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics 225(2), 175–199.

Sundberg, R. and E. Melander (2013). Introducing the UCDP Georeferenced Event Dataset. Journal of Peace Research 50(4), 1989–2019.

Tapsoba, A. (2022, March). Conflict prediction using kernel density estimation. Technical Report 239, AFD Research Papers Series.

Taylor, J., M. Filipski, and M. Alloush (2016). Economic impact of refugees. Proceedings of the National Academy of Sciences 113(27), 7449–53.

Thakral, N. and L. To (2023, July). When Are Estimates Independent of Measurement Units? Mimeo.

Turner, T. (2007). The Congo Wars: Conflict, myth and reality. Zed Books.

United Nations High Commissioner for Refugees (2020). Global Trends: Forced Displacement in 2019. Geneva.

United Nations High Commissioner for Refugees (2023). Global trends. Forced Displacement in 2022. United Nations High Commission for Refugees.

United Nations High Commissioner for Refugees – The UN Refugee Agency (2024). Refugee Data Finder. UNHCR's Refugee Population Statistics Database. Retrieved from https : //www.unhcr.org/ref ugeestatistics/download/?url = R1xq 11..

Vail, L. (1989). The Creation of Tribalism in Southern Africa. Berkeley, CA: University of California Press.

Verme, P. and K. Schuettler (2021). The impact of forced displacement on host communities a review of the empirical literature in economics. Journal of Development Economics 150, 102606.

Verwimp, P. and J. Maystadt (2015). Forced Displacement and Refugees in Sub-Saharan Africa. An Economic Inquiry. World Bank Policy Research Working Paper 7517.

Vogt, Manuel, N.-C. B. S. R. L.-E. C. P. H. and L. Girardin (2015). Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family. Journal of Conflict Resolution 59(7), 1327–42.

Whitaker, B. E. (2003). Refugees and the Spread of Conflict: Contrasting Cases in Central Africa. Journal of Asian and African Studies 38(2-3), 211–231.

Zhou, Y.-Y. (2018). Refugee Proximity and Support for Citizenship Exclusion in Africa. Unpublished.

Zhou, Y.-Y. (2019). How Refugee Resentment Shapes National Identity and Citizen Participation in Africa. Unpublished.

Zhou, Y.-Y., G. Grossman, and S. Ge (2021). When Refugee Exposure Improves Local Development and Public Goods Provision: Evidence from Uganda. Mimeo.

Zhou, Y.-Y. and A. Shaver (2021). Reexamning the Effect of Refugees on Civil Conflict: A Global Subnational Analysis. Amerian Political Science Review 115(4), 1175–1196.

Zylkin, T. (2019). ppml with negative values. Statalist.

Tables, figures and additional information for:

Ethnic diversity and conflict in Sub–Saharan Africa: Evidence from refugee–hosting areas

November 28, 2024

Abstract

This document contains a set of appendices with supplemental material.

Appendix A Context

The number of refugees worldwide rose from around 10 million in 2010 to 20.4 million by the end of 2019 (United Nations High Commissioner for Refugees, 2020). Although refugees are now traveling greater distances than in the 1980s (Devictor et al., 2021), the majority of refugees are still hosted by neighbouring countries, which often face difficult socio-economic conditions of their own. United Nations High Commissioner for Refugees (2020) estimates that 73 percent of refugees live in neighbouring countries and that developing countries host about 85 percent of the world's refugees.

The number of refugees under the mandate of the United Nations High Commissioner for Refugees (UNHCR) living in sub-Saharan Africa rose from 2.2 to 6.3 million over the same period (United Nations High Commissioner for Refugees, 2020). In other regions, two significant changes occurred more abruptly. The war in Syria led to a significant increase in the number of refugees arriving in Europe and the Middle East (the latter is included in "Asia and Pacific" in the Figure D.6) after 2011. A more recent increase in Latin America has been driven by a surge in Venezuelan refugees. In contrast to these recent events, Africa has seen a steady increase in the number of displaced people between 2005 and 2020. These population movements were largely driven by civil wars and political instability in countries such as South Sudan, the Democratic Republic of Congo, the Central African Republic, Somalia, Burundi, and Eritrea (Figure D.7). Until 2016, the majority of refugees were hosted in neighbouring countries, but we confirm the general trend observed by Devictor et al. (2021) of greater geographical dispersion over time (Figure D.8). Chad, the Democratic Republic of Congo, Ethiopia, Rwanda, South Sudan, Sudan, the United Republic of Tanzania, and Uganda are among the least developed countries hosting the largest numbers of refugees (Figure D.9).

Finally, forced displacement in sub-Saharan Africa is further characterized by the protracted nature of refugee situations (Verwimp and Maystadt, 2015). Figure D.10 shows that the number of protracted refugee situations is not only higher in Africa but has also increased sharply over the past decade.[1]

[1] While recognizing the statistical limitations of such a definition, the United Nations High Commissioner for Refugees (2020, 24) defines a protracted refugee situation as one in which 25,000 or more refugees of the same nationality have been in exile in a given host country for at least five consecutive years (excluding Palestinian refugees under the mandate of the United Nations Relief and Works Agency (UNRWA)).

Appendix B Data

Appendix B.1 Linking Ethnic Data from Africa (LEDA)

LEDA provides an interface - a language tree - to flexibly link ethnic groups from different databases and to calculate the linguistic distances between them. LEDA is currently structured around lists of ethnic groups from 12 original datasets, which are the following:

- Afrobarometer Surveys
- All Minorities at Risk (AMAR)
- Census data from IPUMS
- Ethnic Power Relations (EPR) dataset
- Ethnologue languages
- Political Relevant Ethnic Groups from Posner (2004)
- Ethnic groups in Francois, Trebbi & Rainer (2015)
- Ethnic groups from Fearon (2003)
- GREG Data (based on the Russian Atlas Miradova)
- Demographic and Health Surveys
- Murdock Atlas
- Spatially Interpolated Data on Ethnicity (SIDE)

These lists are structured in the LEDA interface by data source, country, year, or in the case of survey data, survey rounds. In our analysis we use Afrobarometer, EPR, and Murdock Atlas data; therefore we can use LEDA functions to link the different ethnic groups.

LEDA consists of three main types of linkage: binary linkage based on the relationships between sets of language nodes associated with two groups; binary linkage based on linguistic distances; and a full calculation of dyadic linguistic distances.

In our main analysis, we use the second type of linkage, binary linkage based on linguistic distances, and set the level of linkage to “dialect”. This is done using the “mindistlink” function of LEDA, which calculates the minimum linguistic distance between two ethnic groups and thus provides the closest linguistic neighbour for each given ethnic group (see Figure D.4). This function calculates a variable called *distance*, which measures the linguistic distance between two ethnic groups. Mathematically, these distances are calculated as

$$D_{L_1L_2} = 1 - \left(\frac{2d(\omega(L_{1,\dots,O}) \cap \omega(L_{2,\dots,O}))}{d(\omega(L_{1,\dots,O})) + d(\omega(L_{2,\dots,O}))} \right)^{\delta}, \tag{B.1}$$

where $d(\omega(L_{1,\dots,O})$ is the length of the path from the first language to the root of the tree, and $d(\omega(L_{1,\dots,O}) \cap \omega(L_{2,\dots,O})$ is the length of the intersection of the paths from the first and second languages to the root. δ is an exponent to discount distances further from the root of the tree; it is typically set to 0.5.

As a robustness check, we also use the first type of link: binary links based on the relationships between sets of language nodes associated with two groups. This is done with the "setlink" function of LEDA. With this function, the two groups are linked as soon as they share any language node at the level of the language tree specified by the link level.

Specifically, we first use LEDA to obtain linkage tables between the Afrobarometer and EPR data for our main analysis, and the Murdock Atlas and EPR data for our IV strategy, using the "mindistlink" function. We also obtain the same tables using the "setlink" function for robustness. We choose "dialect" as our link level, thus adopting a strict definition of ethnic similarity (vs. difference). We also obtain these tables by choosing "language" as our link level for robustness. We therefore end up with 4 linkage tables. Note that when using the "setlink" function, choosing "dialect" or "language" as our link level yields the same linkage table between the Afrobarometer and EPR data.

As a reminder, we use data on ethnicity from the Afrobarometer to obtain the host country diversity indices (data on the ethnicity of respondents in the host country), while we use data on ethnicity from the EPR dataset to define the revised refugee diversity indices (data on the ethnicity of refugees in camps in the host country). Finally, we use the ethnicity data from the Murdock Atlas to obtain the historical home of the refugees we use in our IV approach.

These linkage tables between the different LEDA databases do not provide one-to-one links. Indeed, as ethnicities are identified at different levels in these different databases, they may be linked to several others. In other words, we still need to find a way to arrive at a single definition of ethnicity in our analysis. Overall, the Afrobarometer and Murdock Atlas ethnicities are defined at a more disaggregated level than the EPR ethnicities. As we can aggregate the disaggregated ethnicities but not the aggregated ones, we rename the Afrobarometer and Murdock Atlas ethnicities based on the EPR ethnicities where possible.

We merge these tables with the ethnicity data we have from rounds 3-6 of the Afrobarometer and the UNHCR refugee camp data for the corresponding period, 2005-2016. We drop all pairs of links between ethnicities if they do not occur simultaneously in the Afrobarometer and UNHCR refugee camp data. In other words, we keep only the data on the association between ethnicities that are present in our database.

We isolate one-to-one (injective) relationships between ethnicities in the Afrobarometer and UNHCR refugee camp data. These are trivial to handle (see Figure D.11).

We also isolate many-to-one (bijective) relationships. In this case, we have to aggregate the Afrobarometer ethnicities with their unique and more aggregated correspondence in the UNHCR refugee camps data (see Figure D.12).

The remaining correspondences are either (i) one-to-many (bijective) but opposite to Figure D.12 (i.e. many ethnicities from the UNHCR refugee camp data correspond to one ethnicity from the Afrobarometer) or (ii) many-to-many relationships. In both cases, we take a more pragmatic approach:

a. In both cases, we disregard ethnicities that do not appear in either the Afrobarometer or the UNHCR refugee camp data. This means that for the remaining ethnicity that has no counterpart in either the Afrobarometer or UNHCR refugee camp data, we simply keep the name of the ethnicity as such, i.e. this information is not dropped.

b. Then, after ignoring ethnicities that do not occur in our datasets, we check whether the one-to-many or many-to-many relationship has not boiled down to a one-to-one or many-to-one relationship again. If so, we can treat it as above.

   the remaining one-to-many relationships, we keep these ethnicities as such in the Afrobarometer and consider them as a single ethnic group. Some manual editing may further improve the correspondence.

   the few remaining many-to-many relationships, we consider the ethnicities on either side as separate ethnicities. Again, some further manual treatment may improve the correspondence.

Appendix B.2 Afrobarometer data

Sampling frame. "The sampling frame normally includes all citizens aged 18 and over. As a standard practice, they [we] exclude people living in institutionalized settings, such as students in halls of residence, patients in hospitals, and people in prisons or nursing homes." (Afrobarometer, 2020). Because the sampling frame is based on recent censuses to represent all citizens of voting age in a given country, Afrobarometer samples are unlikely to include refugees. Sample stratification "reduces the likelihood that distinctive ethnic or linguistic groups will be omitted from the sample. Afrobarometer occasionally deliberately oversamples certain populations that are politically significant within a country to ensure that the sub-sample is large enough to be analyzed." (Afrobarometer, 2020).

Clusters. "Clusters represent classes based on location, including administrative regions (such as states or provinces), populated places (such as cities or villages), structures (such as buildings, bridges, or roads), and other topographical features (such as rivers, mountains or national parks), along with precise or approximate geographic information. They are identified by a precision code that allows the user to select the desired level of geographical unit. The Afrobarometer geocoding methodology involves a double-blind process developed by AidData. Trained geocoders assign latitude/longitude and standardised place names to Enumeration Areas (EAs) using a defined hierarchy of geographic terms. Two independent experts use a double-blind coding system, consulting databases such as Geonames and Google Maps. Disagreements trigger an arbitration round for reconciliation, resulting in a master set of geocodes. The approach captures geographic information at different levels (coordinates, city and administrative divisions). Unique to the Afrobarometer, locations are coded as exact or approximate based on specific criteria, using a hierarchy of place names. Quality assurance includes de-duplication and consistency checks to ensure spatial accuracy within country boundaries. Data quality assessment includes factual accuracy, granularity and availability of higher level information. Spatial distribution and precision codes represent the quality and quantity of geocoded data over time. Precision codes represent levels of location granularity, with lower values indicating greater precision. We restrict our analysis to observations with a maximum precision code of 2, covering locations defined at any level smaller than administrative regions. (BenYishay et al., 2017)

individual data. The following Afrobarometer questions are used as proxies for the individual outcomes:[1]

1 Attack: Over the past year, how often (if ever) have you or anyone in your family: Been physically attacked?

2 Crime: Over the past year, how often (if ever) have you or anyone in your family: Feared crime in your own home?

3 National Identity: Let us suppose that you had to choose between being a [Ghanaian/Kenyan/etc.] and being a [respondent's identity group]. Which of these two groups do you feel most strongly attached to? Ethnic or national identity

4 Protest: Here is a list of actions that people sometimes take as citizens. For each of these, please tell me whether you, personally, have done any of these things during the past year. If not, would you do this if you had the chance: Attended a demonstration or protest march?

5 Theft: Over the past year, how often (if ever) have you or anyone in your family: Had something stolen from your house?

6 General trust: Generally speaking, would you say that most people can be trusted or that you must be very careful in dealing with people?

7 Neighbourhood trust: How much do you trust each of the following types of people: Your neighbours?

8 Institutional trust: How much do you trust each of the following, or haven't you heard enough about them to say: The President/Prime Minister?

Appendix C An Instrumental variable approach

To construct a plausibly exogenous instrumental variable, we first implement a gravity model to predict the number of refugees of a given ethnic group e moving from country o to d at time t, based on the EPR-ER data. Specifically, we estimate the following gravity model

[1] These questions are available in rounds 3-6 of our analysis, except for "General Trust" and "Neighbourhood Trust" which are available in rounds 3 and 5.

$$Ref_{odgt} = \alpha_{od} + \gamma_g + \tau_t + \beta_1 Conflict_{ot-1} + \beta_2 Conflict_{gt-1} + \beta_3 Distance_{gd} + \epsilon_{odgt}, \tag{C.1}$$

where Ref_{odgt} is the stock of refugees of ethnic group g moving from country o to country d in year t. Since we have data on annual refugee stocks and wish to estimate changes in these stocks over time using a gravity model, we include origin-destination fixed effects α_{od} so that identification is based only on changes in stocks over time (Zylkin, 2019).[1] We also include time τ_t and ethnic group fixed effects γ_g. We rely on Murdock's atlas to provide a map of ethnographic regions for Africa and the historical homelands of refugees (Murdock, 1967). To match ethnic groups across datasets, we again use LEDA[2] to link data on ethnicity from Murdock's Atlas with data on ethnicity from the EPR-ER dataset and later with data from Afrobarometer.

We use the sum of conflict events in the historical homeland of ethnic group g in the previous year $t-1$, denoted as $Conflict_{gt-1}$, and we use the mean distance between the historical homeland of ethnic group g and the border of country d to predict the number of refugees of a given ethnic group g moving from country o to d at time t. [3]

To be consistent with the EPR-ER data construction, we restrict our analysis to all origin-destination country pairs that are separated by a maximum distance of $\leq$ 950 km. The predicted numbers of refugees are then converted into predicted shares for the three largest groups to follow the logic of the EPR-ER dataset. We then insert these predicted proportions in the following way:

$$\sum \widehat{Ref_{cgt}} = Ref_{odct} \widehat{Share_{odgt}}, \tag{C.2}$$

where $\widehat{Share_{odgt}} = \frac{Ref_{odgt}}{Ref_{odt}}$ and $Ref_{odt} = \sum_g Ref_{odgt}$.

The predicted (and plausibly exogenous) number of refugees by ethnic group e is then used to construct other (plausibly exogenous) diversity indices to be used as instrumental variables. We use these predicted proportions of refugees per camp c to compute refugee diversity indices, again

[1] In Table D.27 we report results from our gravity model. Column (1) corresponds to equation C.1. Column (2) follows the same specification, except that the dyadic origin-destination fixed effects are replaced by separate origin and destination fixed effects. Conflicts in the country of origin and the distance between the countries of origin and destination have an expected negative effect on the predicted number of refugees. Conflict in the ethnic group's historical homeland and distance to the destination country do not appear to have an impact on this prediction.

[2] More information on LEDA can be found in Appendix B.1.

[3] The construction of the IV follows a long tradition of using the gravity model to predict bilateral migration flows (Ravenstein, 1885, 1989; Crozet, 2004; Mayda, 2010; Garcia et al., 2015; Beine et al., 2016). In our analysis, an important difference comes from the additional dimension introduced by the ethnic group g.

following equations 1 and 2. These indices serve as instrumental variables in the first stage equations corresponding to the 2SLS equivalent of equation 4:

$$REF_{jt} = \alpha_j + \tau_t + \delta_1 \widehat{EF_{jt}} + \delta_2 \widehat{EP_{jt}} + \delta_3 Refuges_{jt} + \delta_5 Q_{jt} + \epsilon_{1,jt} \tag{C.3}$$

and

$$REP_{jt} = \alpha_j + \tau_t + \delta_1 \widehat{EF_{jt}} + \delta_2 \widehat{EP_{jt}} + \delta_3 Refuges_{jt} + \delta_5 Q_{jt} + \epsilon_{2,jt}. \tag{C.4}$$

Appendix D Tables and Figures

Table D.1: Summary Table for Data Availability and Quality for Countries in sub-Saharan Africa

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Data	**The Afrobarometer**							**UNHCR refugee camps**	**EPR-ER**
	Round 1[I]	Round 2[I]	Round 3	Round 4	Round 5	Round 6	Round 7[II]		
Period	1999-2001	2002-2003	2005-2006	2008-2009	2012-2013	2014-2015	2016-2017	2000-2016	1975-2017
Benin			1,198	1,200	1,200	1,200	1,200	Available	Available
Botswana	1,200	1,200	1,200	1,200	1,200	1,200	1,198	Available	Not Available
Burkina Faso				1,200	1,200	1,200	1,200	Available	Available
Burundi					1,200	1,200	1,200	Available	Available
Cape Verde		1,268	1,256	1,264	1,208	1,200	1,202	Not Available	Not Available
Cameroon					1,200	1,182	1,200	Available	Available
Gabon						1,198	1,200	Available	Available
Gambia							2,400	Available	Available
Ghana	2,004	1,200	1,197	1,200	2,400	2,400	1,194	Available	Available
Guinea					1,200	1,200	1,599	Available	Available
Ivory Coast					1,200	1,199	1,200	Available	Available
Kenya		2,398	1,278	1,104	2,399	2,397	1,200	Available	Available
Lesotho	1,177	1,200	1,161	1,200	1,197	1,200	1,200	Not Available	Available
Liberia				1,200	1,199	1,199	1,200	Available	Available
Madagascar			1,350	1,350	1,200	1,200	1,200	Not Available	Not Available
Malawi	1,208	1,200	1,200	1,200	2,407	2,400	1,200	Available	Available
Mali	2,089	1,283	1,244	1,232	1,200	1,200	1,200	Available	Available
Mauritius					1,200	1,200	1,200	Not Available	Not Available
Mozambique		1,400	1,198	1,200	2,400	2,400	1,200	Available	Available
Namibia	1,183	1,199	1,200	1,200	1,200	1,200	1,200	Available	Available
Niger			2,363		1,199	1,200	1,600	Available	Available
Nigeria	3,603	2,428		2,324	2,400	2,400	1,200	Available	Available
Sao Tome and Principe						1,196	1,200	Not Available	Not Available
Senegal		1,200	1,200	1,200	1,200	1,200	1,200	Available	Available
Sierra Leone					1,190	1,191	1,840	Available	Available
South Africa	2,200	2,400	2,400	2,400	2,399	2,390	1,200	Not Available	Available
Sudan[III]					1,199	1,200	2,400	Available	Available
Swaziland					1,200	1,200	1,199	Available	Not Available
Tanzania	2,198	1,223	1,304	1,208	2,400	2,386	1,200	Available	Available
Togo					1,200	1,200	1,199	Available	Available
Uganda	2,271	2,400	2,400	2,431	2,400	2,400	1,200	Available	Available
Zambia	1,198	1,198	1,200	1,200	1,200	1,199	1,200	Available	Available
Zimbabwe	1,200	1,104	1,048	1,200	2,400	2,400	1,200	Available	Available

I There is no data on ethnicity in rounds 1 and 2 of the Afrobarometer.

II There is no geocoded data available for round 7 of the Afrobarometer.

III The question of an individual's ethnicity is not asked in Sudan.

Table D.2: Descriptive Statistics

	(1) Obs.	(2) Mean	(3) Std. Dev.	(4) Min.	(5) Max.	(6) Obs.	(7) Mean	(8) Std. Dev.	(9) Min.	(10) Max.
Conflict Events.	**Panel A: Refugee-Hosting Areas**					**Panel B: Non-Hosting Areas**				
Violent conflict, intensity (IHS, 80km)	2,327	0.8992	1.0807	0	5.6131	12,098	0.9887	1.2984	0	6.5367
Violent conflict, incidence (80km)	2,327	0.5170	0.4998	0	1	12,098	0.4730	0.4993	0	1
Civilian conflict, intensity (IHS)	2,327	0.6370	0.8938	0	4.1591	12,098	0.7649	1.1526	0	6.5309
Civilian conflict, incidence	2,327	0.4083	0.4916	0	1	12,098	0.3944	0.4888	0	1
Non-violent conflict, intensity (IHS, 80km)	2,327	1.2188	1.1382	0	4.8521	12,098	1.4422	1.6081	0	5.7808
Non-violent conflict, incidence (80km)	2,327	0.6162	0.4864	0	1	12,098	0.5640	0.4959	0	1
Protest, intensity (IHS, 80km)	2,327	1.0622	1.1148	0	4.7708	12,098	1.3464	1.5822	0	5.7746
Protest, incidence (80km)	2,327	0.5548	0.4971	0	1	12,098	0.5322	0.4990	0	1
UCDP conflicts, intensity (IHS)	2,327	0.1309	0.5543	0	5.4468	12,098	0.2579	0.7253	0	5.5373
UCDP conflicts, incidence	2,327	0.0812	0.2732	0	1	12,098	0.1422	0.3492	0	1
Diversity Indices										
EF	2,327	0.1182	0.1737	0	0.8125	12,098	0.1594	0.2153	0	0.8337
EP	2,327	0.0478	0.0665	0	0.25	12,098	0.0572	0.0720	0	0.25
REF (Min. Ling. Dist., 80km)	2,327	0.1946	0.1931	0	0.8180	12,098	0.1594	0.2153	0	0.8337
REP (Min. Ling. Dist., 80km)	2,327	0.0718	0.0714	0	0.25	12,098	0.0572	0.0720	0	0.25
REF (no intergroup distance)	2,327	0.3790	0.2446	0	0.8494	12,098	0.2634	0.2632	0	0.8664
REP (no intergroup distance)	2,327	0.1407	0.0782	0	0.25	12,098	0.1008	0.0920	0	0.25
Other variables										
Refugees (80km, IHS)	2,327	6.9173	4.6184	0	13.7611	12,098	0	0	0	0
Rain anomalies (80km)	2,327	1.0191	10.1544	-48.2476	44.6816	12,098	0.0473	10.5546	-57.7804	44.8193
Temperature anomalies (80km)	2,327	0.1001	0.2171	-0.5537	1.2996	12,098	0.1161	0.2209	-0.5938	1.1624

Notes: EF, EP: standard diversity indices. REF (80 km, min. ling. dist.), REP (80 km, min. ling. dist.): revised refugee diversity indices using the "minimum linguistic distance" function of LEDA. Refugees (80 km, IHS): Refugees in camps in an 80 km buffer around each cluster.

Table D.3: Jakiela's diagnostic test for heterogeneous treatment effects

	(1)	(2)	(3)	(4)
	Residualized Violent Conflict, Intensity			
Residualized REF (Min. Ling. Dist., 80km)	-0.4189**	-0.3284*	-0.3051	-0.3208*
	(0.1849)	(0.1859)	(0.1866)	(0.1867)
Residualized REP (Min. Ling. Dist., 80km)	1.6482***	1.5492***	1.5237***	1.5434***
	(0.4994)	(0.4996)	(0.4999)	(0.4998)
Treatment group × Residualized REF	-0.0823	0.1238	-0.0247	-0.0481
	(0.4866)	(0.4886)	(0.4994)	(0.4992)
Treatment group × Residualized REP	-1.7524	-1.8830	-1.7607	-1.7013
	(1.2737)	(1.2732)	(1.2760)	(1.2755)
Residualized Refugees (80km, IHS)		-0.0180***	-0.0226***	-0.0227***
		(0.0041)	(0.0052)	(0.0052)
Treatment group × Residualized Refugees			0.0122	0.0128
			(0.0085)	(0.0085)
Treatment Group	-0.0142	0.0115	0.0028	0.0021
	(0.0149)	(0.0160)	(0.0171)	(0.0171)
Residualized Rain anomalies				0.0016***
				(0.0006)
Residualized Temp anomalies				-0.0962***
				(0.0371)
Observations	14,425	14,425	14,425	14,425
Year FE	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS, reported in Column (6). The dependent variable is the residual from regressing violent conflict intensity (in IHS) on cluster and year fixed effects. The other variables are also residuals from similar regressions. These variables are interacted with an indicator equal to 1 if the unit is treated (refugees are present). *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

Table D.4: Summary Table: Alternative Transformations

	(1)	(2)	(3)	(4)	(5)	(6)
	Violent conflict (80km)					
	Intensity (IHS)	Weighted Rank	Square Root	Cube Root	Quartic Root	Quint Root
Refugee EF (Min. Ling. Dist., 80km)	-0.3114	-0.0717	-0.8469	-0.3959*	-0.3019*	-0.2659*
	(0.2686)	(0.0536)	(0.5522)	(0.2404)	(0.1755)	(0.1521)
Refugee EP (Min. Ling. Dist., 80km)	1.2526*	0.2488*	2.5450*	1.2417**	0.9252**	0.7927**
	(0.6919)	(0.1410)	(1.3708)	(0.6099)	(0.4513)	(0.3934)
Refugees (80km)	-0.0183***	-0.0509*	-0.0017***	-0.0043**	-0.0059	-0.0067
	(0.0062)	(0.0271)	(0.0006)	(0.0020)	(0.0038)	(0.0056)
Rain anomalies (80km)	0.0016**	-0.0001	0.0029***	0.0009	0.0002	-0.0001
	(0.0008)	(0.0002)	(0.0010)	(0.0006)	(0.0005)	(0.0005)
Temp anomalies (80km)	-0.0971**	-0.0356***	0.0130	-0.0686*	-0.0822***	-0.0862***
	(0.0450)	(0.0105)	(0.0626)	(0.0361)	(0.0301)	(0.0278)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.807	0.762	0.792	0.790	0.772	0.755
Elasticity: EF	-0.0679	-0.0305	-0.164	-0.106	-0.0930	-0.0887
Elasticity: EP	0.0927	0.0359	0.168	0.113	0.0968	0.0899

Notes: Estimated equation: Equation (4) using OLS, reported in Column (6). Column (1) corresponds to our benchmark estimation presented in Column (6) of Table 2. In Column (2), violent protests are measured using the rank of this variable. Column (3) to Column (6) measure this same variable - violent protests - taking respectively the square, cube, quartic, and quint roots. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number of refugees in the camps within this distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between the Afrobarometer and EPR-ER datasets. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

* More information on LEDA in Appendix B.1.

Table D.5: Diversity and Violent Conflict, Incidence

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Violent Conflict, Incidence					
Native EF	-0.1243	-0.1251				
	(0.1186)	(0.1189)				
Native EP	0.6103*	0.6126*				
	(0.3178)	(0.3186)				
Refugees (80km, IHS)		0.0004		0.0007		0.0003
		(0.0028)		(0.0029)		(0.0029)
REF (Min. Ling. Dist., 80km)			-0.1747	-0.1803	-0.1932*	-0.1958*
			(0.1113)	(0.1141)	(0.1115)	(0.1145)
REP (Min. Ling. Dist., 80km)			0.4520	0.4573	0.4941*	0.4965*
			(0.2926)	(0.2939)	(0.2938)	(0.2953)
Rain anomalies (80km)					-0.0008**	-0.0008**
					(0.0004)	(0.0004)
Temp anomalies (80km)					-0.0875***	-0.0874***
					(0.0230)	(0.0230)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.662	0.662	0.662	0.662	0.662	0.662
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS and an alternative dependent variable: violent conflict incidence, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row B of Table 3.

* More information on LEDA can be found in Appendix B.1.

Table D.6: Diversity and Non-Violent Conflict, Intensity

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Non-Violent Conflict, Intensity					
Native EF	0.0468	0.0790				
	(0.2909)	(0.2902)				
Native EP	0.0358	-0.0580				
	(0.7807)	(0.7809)				
Refugees (80km, IHS)		-0.0152***		-0.0149**		-0.0148**
		(0.0057)		(0.0058)		(0.0058)
REF (Min. Ling. Dist., 80km)			-0.3133	-0.1968	-0.2932	-0.1785
			(0.2787)	(0.2773)	(0.2770)	(0.2757)
REP (Min. Ling. Dist., 80km)			0.7261	0.6150	0.6888	0.5807
			(0.7289)	(0.7210)	(0.7240)	(0.7164)
Rain anomalies (80km)					-0.0006	-0.0007
					(0.0006)	(0.0006)
Temp anomalies (80km)					0.0997**	0.0955**
					(0.0446)	(0.0444)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.862	0.862	0.862	0.862	0.862	0.862
Year	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS and an alternative dependent variable, the incidence of non-violent conflict, reported in Column (6). Columns (1) and (2) introduce the standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number of refugees in the camps within this distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between the Afrobarometer and EPR-ER datasets. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for revised ethnic fractionalization (REF) and revised ethnic polarization (REP) in Column (6) presented in Row C of Table 3.

[*] More information on LEDA can be found in Appendix B.1.

Table D.7: Diversity and Non-Violent Conflict, Incidence

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Non-Violent Conflict, Incidence					
Native EF	-0.0009	-0.0080				
	(0.1150)	(0.1153)				
Native EP	0.2259	0.2466				
	(0.3189)	(0.3188)				
Refugees (80km, IHS)		0.0034		0.0029		0.0027
		(0.0025)		(0.0026)		(0.0026)
REF (Min. Ling. Dist., 80km)			-0.0791	-0.1021	-0.0724	-0.0931
			(0.1088)	(0.1117)	(0.1076)	(0.1105)
REP (Min. Ling. Dist., 80km)			0.4547	0.4766	0.4516	0.4711
			(0.3009)	(0.3035)	(0.2978)	(0.3002)
Rain anomalies (80km)					-0.0019***	-0.0019***
					(0.0003)	(0.0003)
Temp anomalies (80km)					0.0387*	0.0395**
					(0.0198)	(0.0198)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.704	0.704	0.704	0.704	0.705	0.705
Year	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS and an alternative dependent variable: incidence of non-violent conflict, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row D of Table 3.

[*] More information on LEDA can be found in Appendix B.1.

Table D.8: Diversity and Civilian Conflict, Intensity

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Civilian Conflict (IHS), Intensity					
Native EF	-0.0555	-0.0206				
	(0.2502)	(0.2520)				
Native EP	0.9755	0.8740				
	(0.6493)	(0.6564)				
Refugees (80km, IHS)		-0.0164***		-0.0177***		-0.0178***
		(0.0051)		(0.0053)		(0.0053)
REF (Min. Ling. Dist., 80km)			-0.4675*	-0.3290	-0.4661*	-0.3279
			(0.2396)	(0.2435)	(0.2392)	(0.2432)
REP (Min. Ling. Dist., 80km)			1.6414***	1.5094**	1.6414***	1.5113**
			(0.5999)	(0.6027)	(0.5989)	(0.6019)
Rain anomalies (80km)					-0.0005	-0.0006
					(0.0007)	(0.0007)
Temp anomalies (80km)					0.0084	0.0033
					(0.0391)	(0.0390)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.805	0.805	0.805	0.805	0.805	0.805
Year	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS and an alternative dependent variable: intensity of civilian conflict, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row E of Table 3.

[*] More information on LEDA can be found in Appendix B.1.

Table D.9: Diversity and Civilian Conflict, Incidence

	(1)	(2)	(3)	(4)	(5)	(6)[a]
			Civilian Conflict, Incidence			
Native EF	0.0398	0.0500				
	(0.1117)	(0.1127)				
Native EP	0.2539	0.2243				
	(0.3000)	(0.3035)				
Refugees (80km, IHS)		-0.0048*		-0.0053*		-0.0059**
		(0.0027)		(0.0027)		(0.0027)
REF (Min. Ling. Dist., 80km)			-0.1227	-0.0817	-0.1337	-0.0882
			(0.1083)	(0.1100)	(0.1087)	(0.1105)
REP (Min. Ling. Dist., 80km)			0.4657*	0.4265	0.5021*	0.4592
			(0.2812)	(0.2837)	(0.2820)	(0.2849)
Rain anomalies (80km)					-0.0025***	-0.0026***
					(0.0004)	(0.0004)
Temp anomalies (80km)					-0.0454**	-0.0471**
					(0.0216)	(0.0215)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.675	0.675	0.675	0.675	0.676	0.677
Year	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS and an alternative dependent variable: incidence of civilian conflict, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row F of Table 3.

[*] More information on LEDA can be found in Appendix B.1.

Table D.10: Diversity and Protests, Intensity

	(1)	(2)	(3)	(4)	(5)	(6)[a]
			Protests (IHS), Intensity			
Native EF	0.1223	0.1510				
	(0.2829)	(0.2823)				
Native EP	-0.0391	-0.1226				
	(0.7529)	(0.7531)				
Refugees (80km, IHS)		-0.0135***		-0.0138***		-0.0132**
		(0.0051)		(0.0052)		(0.0052)
REF (Min. Ling. Dist., 80km)			-0.2443	-0.1365	-0.2229	-0.1205
			(0.2709)	(0.2696)	(0.2713)	(0.2702)
REP (Min. Ling. Dist., 80km)			0.7022	0.5993	0.6478	0.5514
			(0.7067)	(0.6999)	(0.7071)	(0.7006)
Rain anomalies (80km)					0.0020***	0.0019***
					(0.0007)	(0.0007)
Temp anomalies (80km)					0.0983**	0.0946**
					(0.0451)	(0.0449)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.854	0.854	0.854	0.854	0.854	0.854
Year	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS and an alternative dependent variable: intensity of protests, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row G of Table 3.

* More information on LEDA can be found in Appendix B.1.

Table D.11: Diversity and Protests, Incidence

	(1)	(2)	(3)	(4)	(5)	(6)[a]
			Protests, Incidence			
Native EF	-0.0105	-0.0205				
	(0.1148)	(0.1152)				
Native EP	0.2870	0.3162				
	(0.3162)	(0.3160)				
Refugees (80km, IHS)		0.0047*		0.0043		0.0042
		(0.0025)		(0.0026)		(0.0026)
REF (Min. Ling. Dist., 80km)			-0.0980	-0.1313	-0.0978	-0.1306
			(0.1084)	(0.1117)	(0.1085)	(0.1118)
REP (Min. Ling. Dist., 80km)			0.5567*	0.5884*	0.5570*	0.5879*
			(0.2990)	(0.3022)	(0.2992)	(0.3024)
Rain anomalies (80km)					-0.0002	-0.0001
					(0.0004)	(0.0004)
Temp anomalies (80km)					0.0018	0.0030
					(0.0211)	(0.0211)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.696	0.696	0.696	0.697	0.696	0.697
Year	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS and an alternative dependent variable: incidence of protests, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. The "minimum linguistic distance" function from LEDA[*] is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row H of Table 3.

[*] More information on LEDA can be found in Appendix B.1.

Table D.12: Diversity and Refugee-related Conflict, Intensity

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Refugee-related Conflict (IHS), Intensity					
Native EF	-0.0569	-0.0492				
	(0.0422)	(0.0410)				
Native EP	0.1572	0.1352				
	(0.1099)	(0.1073)				
Refugees (80km, IHS)		-0.0031*		-0.0039**		-0.0039**
		(0.0017)		(0.0016)		(0.0016)
REF (Min. Ling. Dist., 80km)			-0.0603	-0.0231	-0.0632	-0.0260
			(0.0519)	(0.0479)	(0.0522)	(0.0481)
REP (Min. Ling. Dist., 80km)			0.3401***	0.3011**	0.3448***	0.3062**
			(0.1313)	(0.1285)	(0.1322)	(0.1292)
Rain anomalies (80km)					0.0001	0.0000
					(0.0001)	(0.0001)
Temp anomalies (80km)					-0.0126	-0.0143*
					(0.0082)	(0.0083)
Observations	10,427	10,427	10,427	10,427	10,427	10,427
R-squared	0.336	0.336	0.336	0.337	0.337	0.338
Year	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS and an alternative dependent variable: intensity of refugee-related conflicts, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row I of Table 3.

* More information on LEDA can be found in Appendix B.1.

Table D.13: Diversity and Refugee-related Conflict, Incidence

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Refugee-related Conflict, Incidence					
Native EF	-0.0806	-0.0582				
	(0.1007)	(0.1020)				
Native EP	0.2537	0.1883				
	(0.2829)	(0.2835)				
Refugees (80km, IHS)		-0.0106***		-0.0114***		-0.0114***
		(0.0023)		(0.0023)		(0.0023)
REF (Min. Ling. Dist., 80km)			-0.0880	0.0012	-0.0799	0.0080
			(0.0973)	(0.1021)	(0.0978)	(0.1027)
REP (Min. Ling. Dist., 80km)			0.3353	0.2502	0.3191	0.2363
			(0.2661)	(0.2741)	(0.2673)	(0.2754)
Rain anomalies (80km)					-0.0000	-0.0001
					(0.0003)	(0.0003)
Temp anomalies (80km)					0.0395*	0.0362*
					(0.0214)	(0.0214)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.696	0.696	0.696	0.697	0.696	0.697
Year	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS and an alternative dependent variable: incidence of refugee-related conflicts, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row J of Table 3.

* More information on LEDA can be found in Appendix B.1.

Table D.14: Diversity and UCDP Major Conflicts, Intensity

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	UCDP Major Conflicts (IHS), Intensity					
Native EF	-0.0707	-0.0885				
	(0.2270)	(0.2278)				
Native EP	0.0461	0.0980				
	(0.5764)	(0.5792)				
Refugees (80km, IHS)		0.0084**		0.0092**		0.0095**
		(0.0038)		(0.0041)		(0.0041)
REF (Min. Ling. Dist., 80km)			-0.0631	-0.1352	-0.0482	-0.1219
			(0.2035)	(0.2119)	(0.2023)	(0.2108)
REP (Min. Ling. Dist., 80km)			0.1188	0.1876	0.0865	0.1559
			(0.5062)	(0.5119)	(0.5033)	(0.5089)
Rain anomalies (80km)					0.0004	0.0004
					(0.0005)	(0.0005)
Temp anomalies (80km)					0.0714**	0.0741***
					(0.0282)	(0.0279)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.728	0.728	0.728	0.728	0.728	0.728
Year	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS and an alternative dependent variable: UCDP major conflict intensity, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row K of Table 3.

* More information on LEDA can be found in Appendix B.1.

Table D.15: Diversity and UCDP Major Conflicts, Incidence

	(1)	(2)	(3)	(4)	(5)	(6)[a]
Dependent variable:	UCDP Major Conflicts, Incidence					
Native EF	0.0528	0.0444				
	(0.0823)	(0.0827)				
Native EP	-0.1593	-0.1351				
	(0.2143)	(0.2154)				
Refugees (80km, IHS)		0.0039***		0.0040***		0.0041***
		(0.0013)		(0.0014)		(0.0014)
REF (Min. Ling. Dist., 80km)			0.0473	0.0158	0.0574	0.0255
			(0.0747)	(0.0775)	(0.0742)	(0.0769)
REP (Min. Ling. Dist., 80km)			-0.1039	-0.0739	-0.1229	-0.0928
			(0.1908)	(0.1923)	(0.1891)	(0.1903)
Rain anomalies (80km)					-0.0002	-0.0002
					(0.0003)	(0.0003)
Temp anomalies (80km)					0.0498***	0.0509***
					(0.0152)	(0.0152)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.703	0.703	0.703	0.703	0.703	0.703
Year	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS and an alternative dependent variable: UCDP major conflict incidence, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row L of Table 3.

[*] More information on LEDA can be found in Appendix B.1.

Table D.16: Ethnic Fractionalization and Polarization without Inter-group Distance

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Violent Conflict, Incidence					
EF (no dist.)	0.0940	0.1215				
	(0.1954)	(0.1959)				
EP (no dist.)	-0.1074	-0.1829				
	(0.5490)	(0.5502)				
Refugees (80km, IHS)		-0.0180***		-0.0170***		-0.0171***
		(0.0058)		(0.0065)		(0.0065)
REF (Min. Ling. Dist., 80km) (no dist.)			-0.4397**	-0.3029	-0.4470**	-0.3105
			(0.1980)	(0.2072)	(0.1995)	(0.2087)
REP (Min. Ling. Dist., 80km) (no dist.)			0.9780*	0.8947*	1.0042*	0.9221*
			(0.5134)	(0.5152)	(0.5171)	(0.5186)
Rain anomalies (80km)					0.0017**	0.0016**
					(0.0008)	(0.0008)
Temp anomalies (80km)					-0.0931**	-0.0981**
					(0.0451)	(0.0450)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.807	0.807	0.807	0.807	0.807	0.807
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Revised refugee diversity indices are not weighted by intergroup distance. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row B of Table 4.

[*] More information on LEDA can be found in Appendix B.1.

Table D.17: Ethnic Fractionalization without Inter-group Distance

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Violent Conflict, Incidence					
EF (no dist.)	-0.1147	-0.1008				
	(0.1256)	(0.1255)				
EP	1.0532**	1.0022**				
	(0.4515)	(0.4516)				
Refugees (80km, IHS)		-0.0175***		-0.0164**		-0.0165**
		(0.0059)		(0.0065)		(0.0065)
REF (Min. Ling. Dist., 80km) (no dist.)			-0.3065***	-0.1807	-0.2998**	-0.1735
			(0.1177)	(0.1263)	(0.1182)	(0.1270)
REP (Min. Ling. Dist., 80km)			1.0473**	0.9337**	1.0166**	0.9029**
			(0.4150)	(0.4162)	(0.4171)	(0.4184)
Rain anomalies (80km)					0.0016**	0.0016**
					(0.0008)	(0.0008)
Temp anomalies (80km)					-0.0894**	-0.0946**
					(0.0450)	(0.0450)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.807	0.807	0.807	0.807	0.807	0.807
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. REF is not weighted by intergroup distance while REP is weighted by intergroup distance. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row C of Table 4.

[*] More information on LEDA can be found in Appendix B.1.

Table D.18: Controlling for Ethnic Fractionalization without Inter-group Distance

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Violent Conflict, Incidence					
EF	-0.0437	-0.0148				
	(0.3156)	(0.3190)				
EP	1.1331	1.0292				
	(0.7453)	(0.7518)				
EF (no dist.)	-0.1039	-0.0972				
	(0.1440)	(0.1442)				
Refugees (80km, IHS)		-0.0175***		-0.0166**		-0.0167***
		(0.0059)		(0.0064)		(0.0064)
Refugee EF (Min. Ling. Dist., 80km) (no dist.)			-0.2890**	-0.1480	-0.2713*	-0.1289
			(0.1406)	(0.1475)	(0.1408)	(0.1479)
Refugee EF (Min. Ling. Dist., 80km)			-0.0640	-0.1161	-0.1042	-0.1579
			(0.3098)	(0.3111)	(0.3116)	(0.3127)
Refugee EP (Min. Ling. Dist., 80km)			1.1579*	1.1332*	1.1964*	1.1741*
			(0.6817)	(0.6856)	(0.6877)	(0.6911)
Rain anomalies (80km)					0.0016**	0.0016**
					(0.0008)	(0.0008)
Temp anomalies (80km)					-0.0903**	-0.0959**
					(0.0450)	(0.0449)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.807	0.807	0.807	0.807	0.807	0.807
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. REF is not weighted by distance while REP is weighted by distance. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row D of Table 4.

[*] More information on LEDA can be found in Appendix B.1.

Table D.19: Refugee Camps in a 40-km Buffer Around Clusters

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Violent Conflict, Incidence					
EF	-0.1794	-0.1650				
	(0.2195)	(0.2213)				
EP	0.7634	0.7246				
	(0.5990)	(0.6030)				
Refugees (40km, IHS)		-0.0191***		-0.0190***		-0.0188***
		(0.0059)		(0.0060)		(0.0060)
REF (Min. Ling. Dist., 40km)			-0.3903*	-0.3011	-0.3875*	-0.2999
			(0.2168)	(0.2189)	(0.2167)	(0.2188)
REP (Min. Ling. Dist., 40km)			1.1855**	1.0758*	1.1711**	1.0642*
			(0.5801)	(0.5826)	(0.5799)	(0.5823)
Rain anomalies (40km)					0.0013**	0.0012**
					(0.0006)	(0.0006)
Temp anomalies (40km)					0.0087	0.0066
					(0.0375)	(0.0376)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.803	0.803	0.803	0.803	0.803	0.804
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS, reported in Column (6). Refugee camps in a 40-km buffer around each cluster. Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row E of Table 4.

* More information on LEDA can be found in Appendix B.1.

Table D.20: Refugee Camps in a 120-km Buffer Around Clusters

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Violent Conflict, Incidence					
EF	-0.2916	-0.2867				
	(0.3034)	(0.3039)				
EP	1.4510*	1.4406*				
	(0.8206)	(0.8216)				
Refugees (120km, IHS)		-0.0025		-0.0020		-0.0024
		(0.0052)		(0.0055)		(0.0055)
REF (Min. Ling. Dist., 120km)			-0.4693*	-0.4516	-0.5079*	-0.4867*
			(0.2701)	(0.2769)	(0.2731)	(0.2801)
REP (Min. Ling. Dist., 120km)			1.3256*	1.3151*	1.3962**	1.3839*
			(0.7026)	(0.7041)	(0.7112)	(0.7127)
Rain anomalies (120km)					0.0012	0.0012
					(0.0009)	(0.0009)
Temp anomalies (120km)					-0.1828***	-0.1838***
					(0.0492)	(0.0491)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.815	0.815	0.815	0.815	0.815	0.815
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS, reported in Column (6). Refugee camps in a 120-km buffer around each cluster. Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row F of Table 4.

[*] More information on LEDA can be found in Appendix B.1.

Table D.21: Aggregation at the GADM2 level.

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Violent Conflict, Incidence					
EF	-1.6132*	-1.6155*				
	(0.8412)	(0.8424)				
EP	3.2141	3.2633				
	(2.3503)	(2.3532)				
Refugees (GADM2, IHS)		0.0424		0.0454		0.0478
		(0.0286)		(0.0298)		(0.0299)
REF (Min. Ling. Dist., 80km)			-1.4917*	-1.5113*	-1.5658**	-1.5886**
			(0.7893)	(0.7921)	(0.7966)	(0.7993)
REP (Min. Ling. Dist., 80km)			3.9409*	3.9516*	4.0140*	4.0304*
			(2.1908)	(2.1970)	(2.1867)	(2.1914)
Rain anomalies					0.0056	0.0055
					(0.0065)	(0.0065)
Temp anomalies					0.2309	0.2477
					(0.2792)	(0.2824)
Observations	1,563	1,563	1,563	1,563	1,563	1,563
R-squared	0.811	0.811	0.811	0.811	0.811	0.811
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: GADM2. Number of countries: 23. Period: 2005-2016. Refugee camps in an 80-km buffer around each cluster. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row G of Table 4.

* More information on LEDA can be found in Appendix B.1.

Table D.22: Including all geocoded locations

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Violent Conflict, Incidence					
EF	-0.0963	-0.0724				
	(0.1993)	(0.2001)				
EP	0.6881	0.6261				
	(0.5155)	(0.5179)				
Refugees (80km, IHS)		-0.0175***		-0.0187***		-0.0188***
		(0.0043)		(0.0045)		(0.0045)
REF (Min. Ling. Dist., 80km)			-0.2273	-0.1274	-0.2330	-0.1327
			(0.1845)	(0.1873)	(0.1850)	(0.1878)
REP (Min. Ling. Dist., 80km)			0.7383	0.6959	0.7512	0.7087
			(0.4704)	(0.4716)	(0.4717)	(0.4728)
Rain anomalies (80km)					-0.0002	-0.0002
					(0.0006)	(0.0006)
Temp anomalies (80km)					-0.0586*	-0.0601*
					(0.0353)	(0.0353)
Observations	23,236	23,236	23,236	23,236	23,236	23,236
R-squared	0.787	0.788	0.787	0.788	0.787	0.788
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS, reported in Column (6). All geocoded locations are included. Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. Refugee camps in an 80-km buffer around each cluster. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row H of Table 4.

* More information on LEDA can be found in Appendix B.1.

Table D.23: Using a Non-Linear Model

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Violent Conflict, Incidence					
EF	-0.1203	-0.1250				
	(0.1333)	(0.1319)				
EP	0.6276*	0.6351*				
	(0.3497)	(0.3458)				
Refugees (80km, IHS)		0.0014		0.0021		0.0017
		(0.0018)		(0.0020)		(0.0022)
REF (Min. Ling. Dist., 80km)			-0.1840	-0.2184*	-0.2215	-0.2478*
			(0.1271)	(0.1311)	(0.1402)	(0.1441)
REP (Min. Ling. Dist., 80km)			0.4973	0.5395	0.5843	0.6157*
			(0.3282)	(0.3282)	(0.3613)	(0.3610)
Rain anomalies (80km)					-0.0008**	-0.0008*
					(0.0004)	(0.0004)
Temp anomalies (80km)					-0.0818**	-0.0794**
					(0.0338)	(0.0335)
Observations	5,749	5,749	5,749	5,749	5,749	5,749
Number of cluster_id	1,829	1,829	1,829	1,829	1,829	1,829
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using logit presented in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. An 80-km buffer around each cluster is used to *revise* standard ethnic diversity measures with the number and ethnic composition of refugees in the camps within this distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, **, and *** indicate significance at the 10%, 5% and 1% levels, respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row I of Table 4.

[*] More information on LEDA can be found in Appendix B.1.

Table D.24: Accounting for Historical Ethnic Polarization

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Violent Conflict, Incidence					
EF	-0.1752	-0.1441				
	(0.2750)	(0.2775)				
EP	3.4418***	2.7162***				
	(0.9132)	(0.9705)				
Refugees (80km, IHS)		-0.0139**		-0.0198***		-0.0195***
		(0.0064)		(0.0061)		(0.0061)
REF (Min. Ling. Dist., 80km)			-0.4186	-0.2693	-0.4340*	-0.2874
			(0.2618)	(0.2684)	(0.2635)	(0.2700)
REP (Min. Ling. Dist., 80km)			2.3175**	2.5794***	2.2450**	2.5072**
			(0.9267)	(0.9823)	(0.9363)	(0.9912)
Rain anomalies (80km)					0.0018**	0.0016**
					(0.0008)	(0.0008)
Temp anomalies (80km)					-0.1076**	-0.1117**
					(0.0452)	(0.0452)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.808	0.808	0.807	0.808	0.808	0.808
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y
EP using Murdock Borders*Time FE (historical ethnic polarization)	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS, reported in Column (6). Includes EP using Murdock Borders*time fixed effects to account for historical ethnic polarization. Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. Refugee camps in an 80 km buffer around each cluster. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row J of Table 4.

* More information on LEDA can be found in Appendix B.1.

Table D.25: Accounting for Conflict Spillovers

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Violent Conflict, Incidence					
EF	-0.1657	-0.1288				
	(0.2795)	(0.2823)				
EP	1.2241	1.1172				
	(0.7596)	(0.7662)				
Refugees (80km, IHS)		-0.0183***		-0.0192***		-0.0192***
		(0.0059)		(0.0062)		(0.0062)
REF (Min. Ling. Dist., 80km)			-0.4357*	-0.2864	-0.4514*	-0.3036
			(0.2645)	(0.2713)	(0.2661)	(0.2728)
REP (Min. Ling. Dist., 80km)			1.3515*	1.2096*	1.3743*	1.2360*
			(0.6970)	(0.7016)	(0.7022)	(0.7062)
Rain anomalies (80km)					0.0017**	0.0016**
					(0.0008)	(0.0008)
Temp anomalies (80km)					-0.0952**	-0.1011**
					(0.0465)	(0.0465)
Observations	14,425	14,425	14,425	14,425	14,425	14,425
R-squared	0.808	0.808	0.808	0.808	0.808	0.808
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y
Distance*Time FE (conflict spillovers)	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS, reported in Column (6). Includes distance*time fixed effects to account for conflict spillovers. Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, revised refugee diversity indices are introduced. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. Refugee camps in an 80 km buffer around each cluster. The “minimum linguistic distance” function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Row K of Table 4.

* More information on LEDA can be found in Appendix B.1.

Table D.26: Instrumenting Refugee Location

	(1)	(2)	(3)	(4)[a]
	Violent Conflict (IHS), Intensity			
Panel A:	**Second Stage**			
REF (80 km, min. ling. dist.)	-0.4942*	-0.4589*	-0.3082	-0.2815
	(0.2775)	(0.2755)	(0.2833)	(0.2816)
REP (80 km, min. ling. dist.)	1.5630**	1.5094**	1.3936*	1.3493*
	(0.7355)	(0.7330)	(0.7379)	(0.7357)
Refugees (80 km, IHS)			-0.0264***	-0.0254***
			(0.0073)	(0.0072)
Observations	11,909	11,909	11,909	11,909
R-squared	0.0024	0.0076	0.0058	0.0106
Kleibergen-Paap rk Wald F	3506	3493	3211	3210
Root MSE	0.566	0.564	0.565	0.563
Panel B:	**First Stage (REF)**			
Predicted REF	0.9687***	0.9688***	0.9595***	0.9600***
	(0.0109)	(0.0114)	(0.0120)	(0.0123)
Predicted REP	0.0560*	0.0562	0.0667*	0.0663*
	(0.0337)	(0.0342)	(0.0341)	(0.0345)
Panel C:	**First Stage (REP)**			
Predicted REF	0.0101***	0.0104***	0.0061**	0.0065**
	(0.0029)	(0.0030)	(0.0029)	(0.0029)
Predicted REP	0.9671***	0.9666***	0.9717***	0.9711***
	(0.0078)	(0.0079)	(0.0076)	(0.0078)
Year FE	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y
Conflict Spillovers	N	Y	N	Y
Refugees (80 km, IHS)	N	N	Y	Y
Climatic controls	Y	Y	Y	Y

Notes: Estimated equation in Panel A: Equation (C.1). Estimated equation in Panel B: Equation (C.3). Estimated equation in Panel C: Equation (C.4). Refugee camps in an 80 km buffer around each cluster. The "minimum linguistic distance" function from LEDA is used. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects. REF and REP predicted using a gravity model presented in Column (2) of Table D.27.

[a] Results for REF and REP in Column (4) and Panel A presented in Row L of Table 4.

[*] More information on LEDA can be found in Appendix B.1.

Table D.27: Instrumental Variable Approach: Gravity Model

	(1)	(2)
	Stock of refugees per ethnic group	
Conflict events at origin	0.0008***	0.0008***
	(0.0003)	(0.0003)
Distance, origin-destination	-	-0.0034***
	-	(0.0011)
Conflict events in hist. ethnic homeland	-0.0002	-0.0002
	(0.0002)	(0.0002)
Distance, hist. ethnic homeland-destination	-0.0001	-0.0014**
	(0.0005)	(0.0007)
Destination FE	N	Y
Ethnic Group FE	Y	Y
Origin FE	N	Y
Origin-Destination FE	Y	N
Year FE	Y	Y
Observations	4,068	4,140
Pseudo R-squared	0.667	0.607

Notes: Estimated equation: Equation (C.1) with PPML presented in Column (1). Equation (C.1) with separate origin and destination fixed effects instead of dyadic origin-destination fixed effects, presented in Column (2). Number of countries: 23. Period: 2005-2016. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between the Murdock Atlas and the EPR-ER data. Robust standard errors clustered at the origin and destination are shown in parentheses. FE: Fixed effects.

Table D.28: Descriptive Statistics: Individual Data

	(1) Obs.	(2) Mean	(3) Std. Dev.	(4) Min.	(5) Max.	(6) Obs.	(7) Mean	(8) Std. Dev.	(9) Min.	(10) Max.
	Panel A: Refugee-Hosting Areas					**Panel B: All Areas**				
Diversity Indices										
EF	8,767	0.1276	0.1754	0	0.8028	56,700	0.1738	0.2231	0	0.8337
EP	8,767	0.0511	0.0671	0	0.2500	56,700	0.0594	0.0704	0	0.2500
REF (80 km, min. ling. dist.)	8,767	0.1999	0.1915	0	0.8180	56,700	0.1849	0.2244	0	0.8337
REP (80 km, min. ling. dist.)	8,767	0.0724	0.0700	0	0.2500	56,700	0.0627	0.0709	0	0.2500
Refugees (80 km, IHS)	8,767	6.8401	4.6448	0	13.7611	56,700	1.0576	3.0743	0	13.7611
Sociodemographic Variables										
Age	8,767	36.7967	14.0385	18	100	56,700	36.1851	14.0944	18	130
Basic education	8,767	0.3458	0.4757	0	1	56,700	0.2781	0.4481	0	1
Secondary education	8,767	0.3209	0.4668	0	1	56,700	0.3418	0.4743	0	1
Tertiary education	8,767	0.0806	0.2723	0	1	56,700	0.1214	0.3266	0	1
Female	8,767	0.5009	0.5000	0	1	56,700	0.5005	0.5000	0	1
Marital status	8,767	0.0541	0.2262	0	1	56,700	0.0639	0.2446	0	1
Outcome Variables										
Attacks	8,767	0.0859	0.2802	0	1	56,700	0.1039	0.3052	0	1
Crime	8,767	0.2917	0.4546	0	1	56,700	0.3122	0.4634	0	1
Identity: Ethnicity vs. Nationality	8,767	0.5161	0.4998	0	1	56,700	0.4737	0.4993	0	1
Protest	8,767	0.2480	0.4319	0	1	56,700	0.3280	0.4695	0	1
Theft	8,767	0.3002	0.4584	0	1	56,700	0.3047	0.4603	0	1
Trust: general	4,912	0.2239	0.4169	0	1	27,127	0.2054	0.4040	0	1
Trust: government	8,767	0.6123	0.4873	0	1	56,700	0.6119	0.4873	0	1
Trust: neighbourhood	4,912	0.6154	0.4865	0	1	27,127	0.6304	0.4827	0	1
Climate Data										
Rain anomalies	8,767	-0.0628	11.3990	-48.2476	28.5457	56,700	-0.4404	11.6164	-57.7804	41.6399
Temperature anomalies	8,767	0.0738	0.2395	-0.5414	1.2996	56,700	0.0857	0.2516	-0.5938	1.2996

Notes: EF, EP: standard diversity indices. REF (80 km, min. ling. dist.), REP (80 km, min. ling. dist.): revised refugee diversity indices. Level of analysis: cluster. Number of countries: 23. Period: 2005-2016. Refugee camps in an 80 km buffer around each cluster. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer surveys and EPR-ER data.

* More information on LEDA can be found in Appendix B.1.

Table D.29: Discussion: Diversity and Individual Outcomes

	(1) Attack[a]	(2) Crime	(3) Gen. Trust	(4) Gov. Trust	(5) National Id.	(6) Neigh. Trust	(7) Protest	(8) Theft
REF (80 km, min. ling. dist.)	-0.0762	-0.1497	-0.1384	0.0714	-0.0662	-0.2880	-0.1157*	-0.1149
	(0.0524)	(0.1013)	(0.1958)	(0.0921)	(0.1220)	(0.1914)	(0.0694)	(0.0877)
REP (80 km, min. ling. dist.)	0.3162**	0.3142	0.1978	0.0228	0.1069	0.4818	0.2181	0.1169
	(0.1584)	(0.2789)	(0.5207)	(0.2510)	(0.3358)	(0.5988)	(0.2011)	(0.2493)
Refugees (80 km, IHS)	0.0009	0.0020	-0.0030	0.0035	0.0035	-0.0005	-0.0027	0.0012
	(0.0015)	(0.0042)	(0.0071)	(0.0033)	(0.0045)	(0.0110)	(0.0033)	(0.0039)
Observations	56,700	56,700	27,126	56,700	56,700	27,126	56,700	56,700
R-squared	0.160	0.195	0.229	0.263	0.225	0.295	0.496	0.175
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y	Y	Y
Controls: climate, Ind.	Y	Y	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (5) using OLS. Individual controls: age, age squared, education, sex, marital status. and rural/urban status. Level of analysis: cluster. Period: 2005-2016. Refugee camps in an 80 km buffer around each cluster. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between Afrobarometer and EPR-ER data. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects. See Column (6) for: [a] in Table D.30.

* More information on LEDA can be found in Appendix B.1.

Table D.30: Diversity and Attacks

	(1)	(2)	(3)	(4)	(5)	(6)[a]
	Attack					
EF	-0.0548	-0.0588				
	(0.0550)	(0.0549)				
EP	0.2636	0.2735*				
	(0.1615)	(0.1613)				
Refugees (80 km, IHS)		0.0013		0.0013		0.0009
		(0.0014)		(0.0015)		(0.0015)
REF (80 km, min. ling. dist.)			-0.0630	-0.0706	-0.0713	-0.0762
			(0.0525)	(0.0533)	(0.0514)	(0.0524)
REP (80 km, min. ling. dist.)			0.2927*	0.3059*	0.3078*	0.3162**
			(0.1587)	(0.1582)	(0.1587)	(0.1584)
Rain anomalies					0.0005	0.0005
					(0.0004)	(0.0004)
Temp. anomalies					-0.0388	-0.0378
					(0.0246)	(0.0249)
Observations	56,700	56,700	56,700	56,700	56,700	56,700
R-squared	0.159	0.159	0.159	0.159	0.159	0.160
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y
Individual Controls	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (5) using OLS, reported in Column (6). Columns (1) and (2) introduce standard diversity indices. From Column (3) onwards, the revised refugee diversity indices are presented. Individual controls: age, age squared, education, sex, marital status, and rural/urban status. Refugee camps in an 80 km buffer around each cluster. The "minimum linguistic distance" function from LEDA is used. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

[a] Results for REF and REP in Column (6) presented in Column (1) of Table D.29.

[*] More information on LEDA can be found in Appendix B.1.

Table D.31: Diversity and Violent Conflict without Uganda

	(1)	(2)	(3)	(4)	(5)	(6)
			Violent Conflict, Intensity			
Native EF	-0.1661	-0.1289				
	(0.2769)	(0.2790)				
Native EP	1.1580	1.0489				
	(0.7475)	(0.7533)				
Refugees (km)		-0.0176***		-0.0184***		-0.0180***
		(0.0060)		(0.0063)		(0.0062)
Refugee EF (Min. Ling. Dist., 80km)			-0.4509*	-0.3053	-0.4670*	-0.3256
			(0.2620)	(0.2687)	(0.2651)	(0.2717)
Refugee EP (Min. Ling. Dist., 80km)			1.3548**	1.2161*	1.3682**	1.2359*
			(0.6862)	(0.6905)	(0.6945)	(0.6982)
Rain anomalies (80km)					0.0034***	0.0033***
					(0.0008)	(0.0008)
Temp anomalies (80km)					-0.0931**	-0.0995**
					(0.0458)	(0.0457)
Constant	0.8975***	0.9164***	0.9310***	0.9342***	0.9424***	0.9462***
	(0.0208)	(0.0220)	(0.0203)	(0.0202)	(0.0210)	(0.0208)
Observations	13,798	13,798	13,798	13,798	13,798	13,798
R-squared	0.806	0.806	0.806	0.806	0.807	0.807
Year FE	Y	Y	Y	Y	Y	Y
PSU FE	Y	Y	Y	Y	Y	Y

Notes: Estimated equation: Equation (4) using OLS, presented in Column (6) excluding Uganda. Columns (1) and (2) present the standard diversity indices. From Column (3) the revised refugee diversity indices are presented. Level of analysis: cluster. Number of countries: 22. Period: 2005-2016. An 80 km buffer around each cluster is used to *revise* the standard ethnic diversity measures with the number of refugees in camps within that distance. The "minimum linguistic distance" function from LEDA* is used to link ethnicities between the Afrobarometer and EPR-ER datasets. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. Robust standard errors clustered at the cluster level are shown in parentheses. FE: fixed effects.

* More information on LEDA in Appendix B.1.

Figure D.1: UNHCR Official Refugee Statistics vs. UNHCR Refugee Camps: Aggregated Data

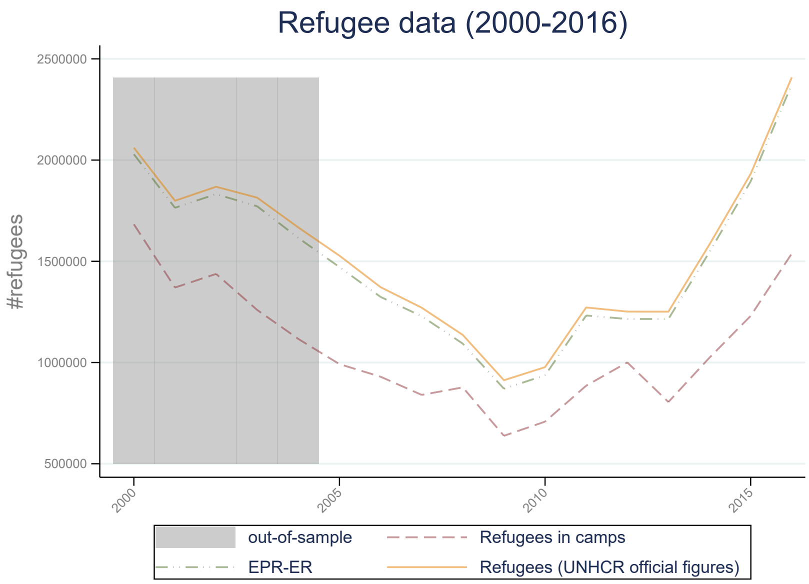


Figure D.2: Refugee Flows between Source and Asylum Countries

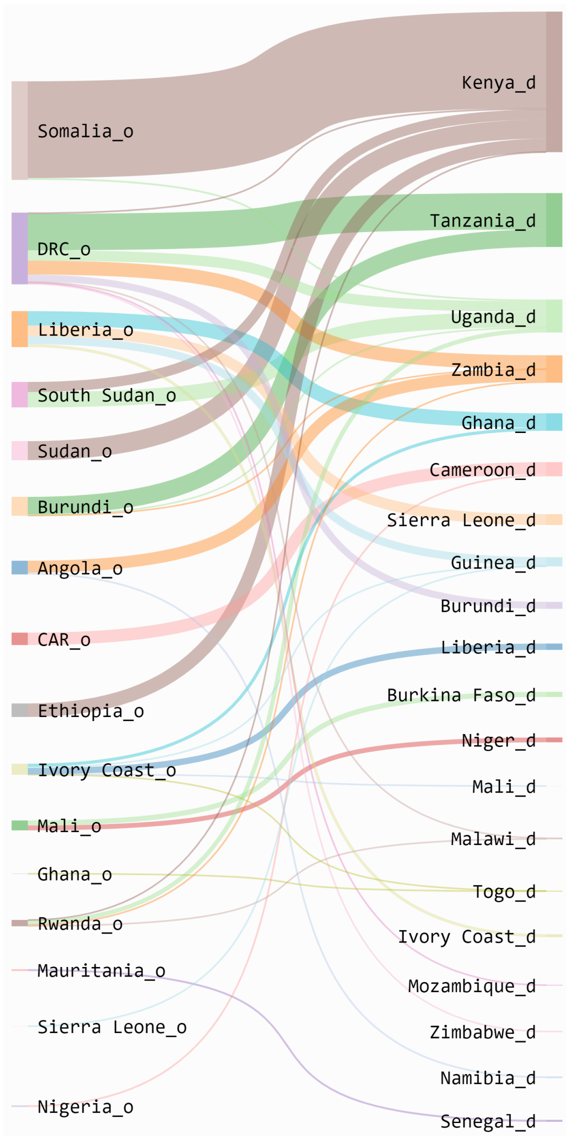
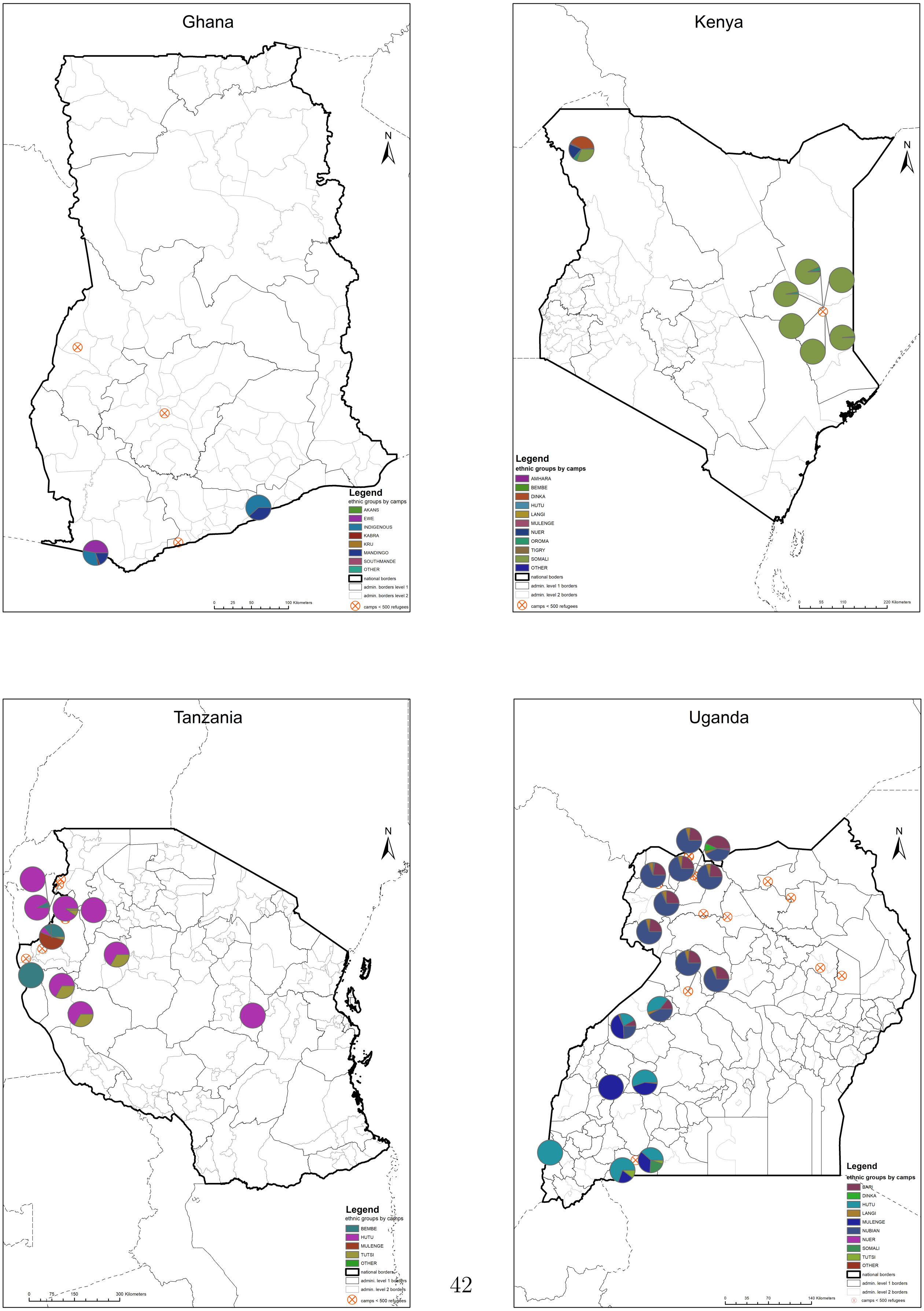


Figure D.3: Refugees in Camps per Ethnic Group



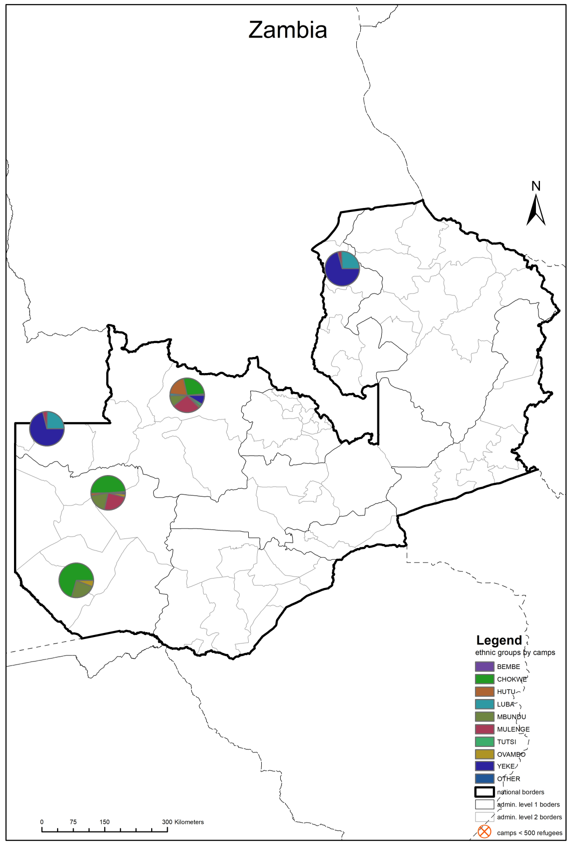
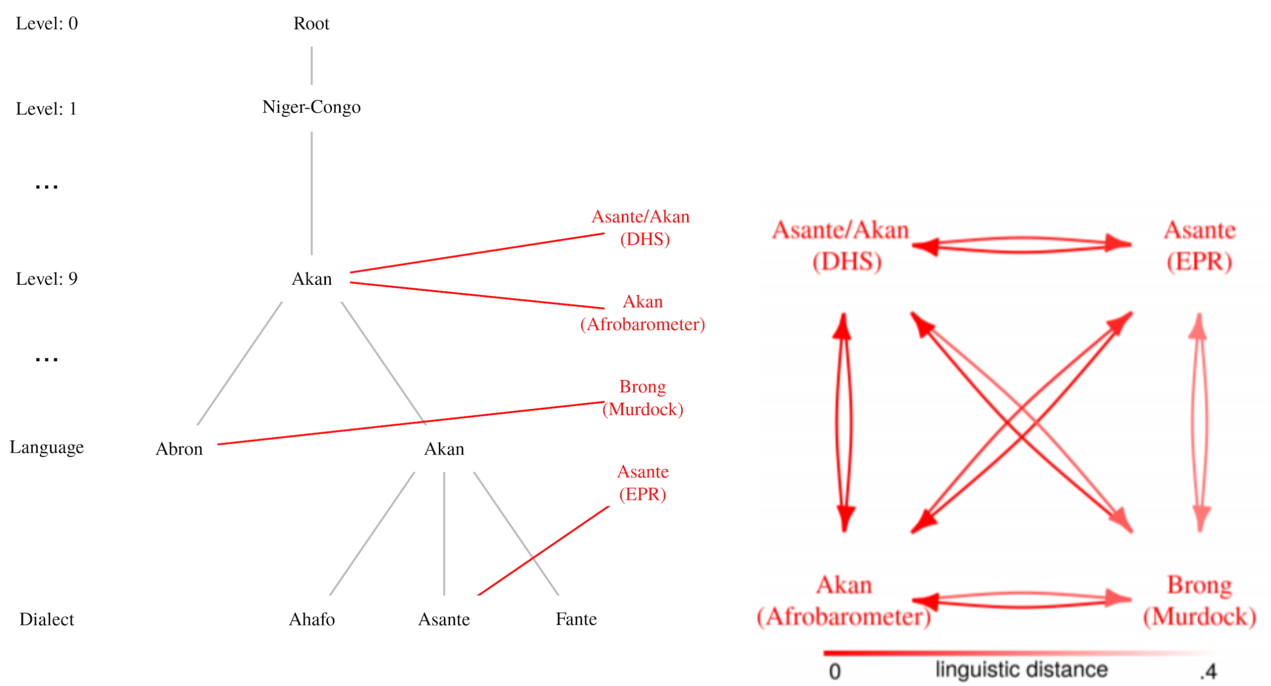

Zambia
N
Legend
ethnic groups by camps
BEMBE
CHOKWE
HUTU
LUBA
MBUNDU
MULENGE
TUTSI
OVAMBO
YEKE
OTHER
national borders
admin. level 1 boders
admin. level 2 borders
camps < 500 refugees
0 75 150 300 Kilometers

Figure D.4: Linking Ethnic Data from Africa



Source: Müller-Crepon et al., 2020.

Figure D.5: Heterogeneity Analysis

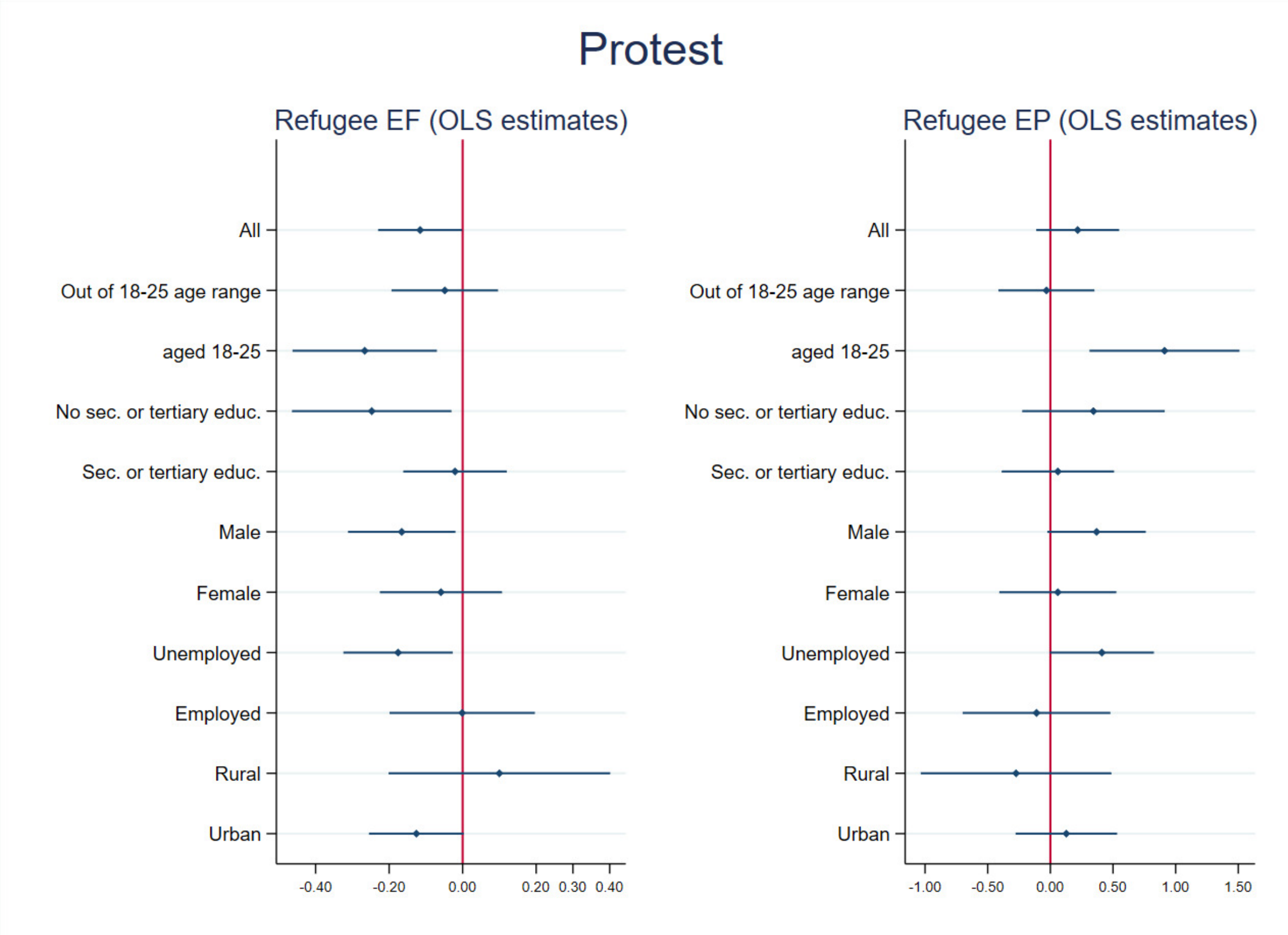


Figure D.6: UNHCR Aggregated Refugee Data by Region of Asylum

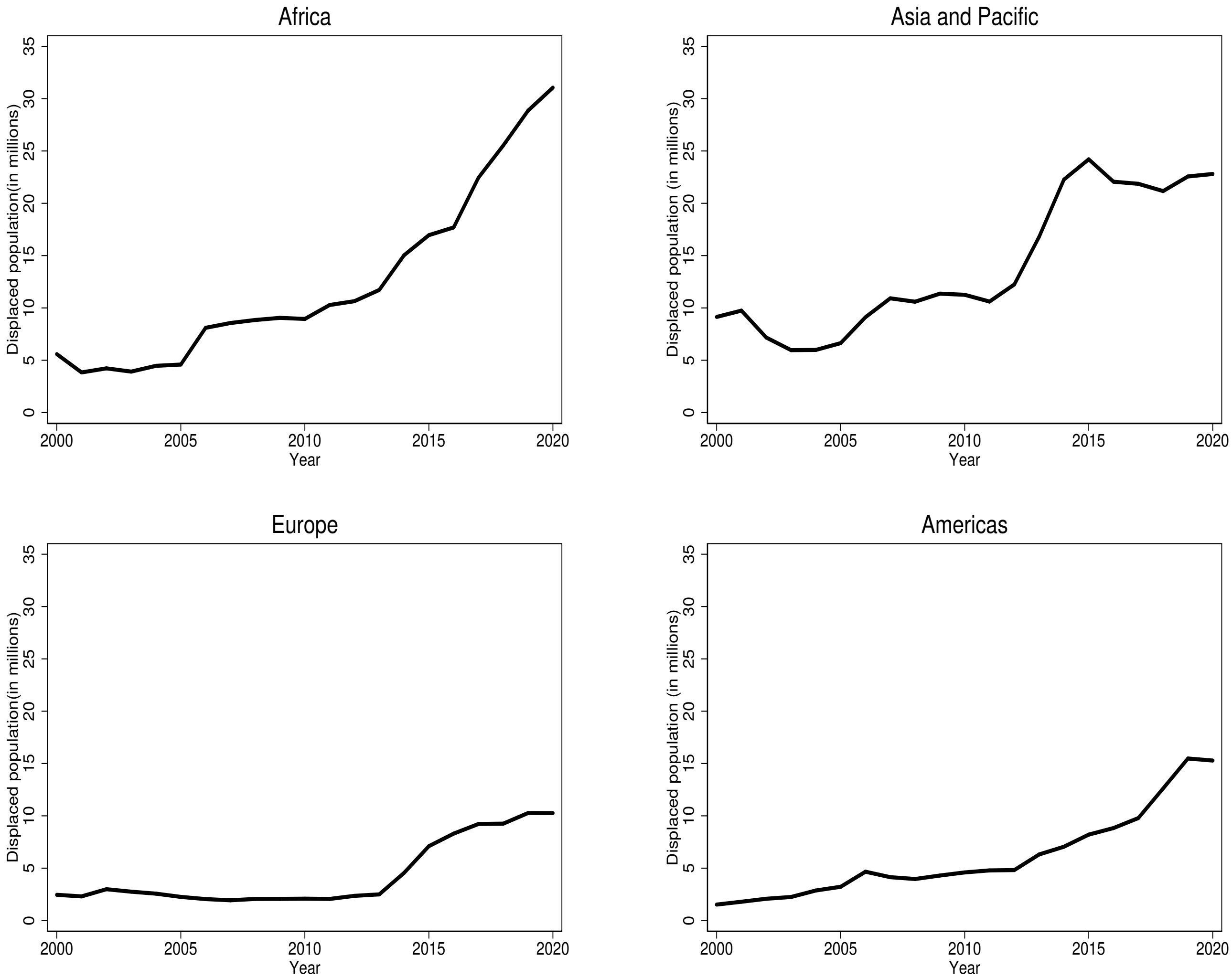
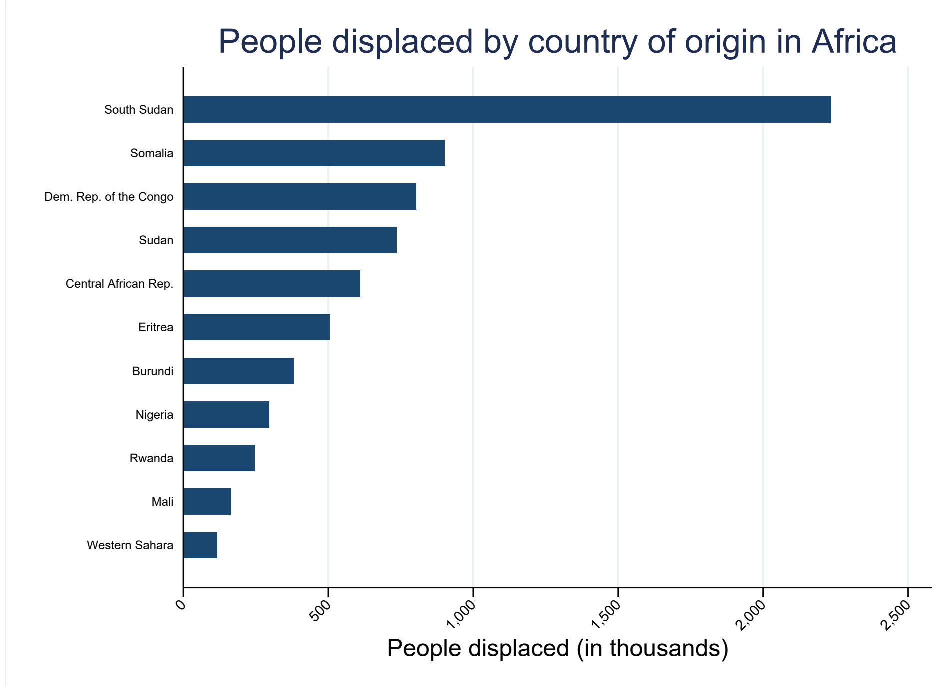


Figure D.7: People Displaced Across Borders by Country of Origin, UNHCR, End of 2019



Note: Countries with more than 100,000 refugees at the origin.

Figure D.8: UNHCR Share of Refugees in neighbouring Countries

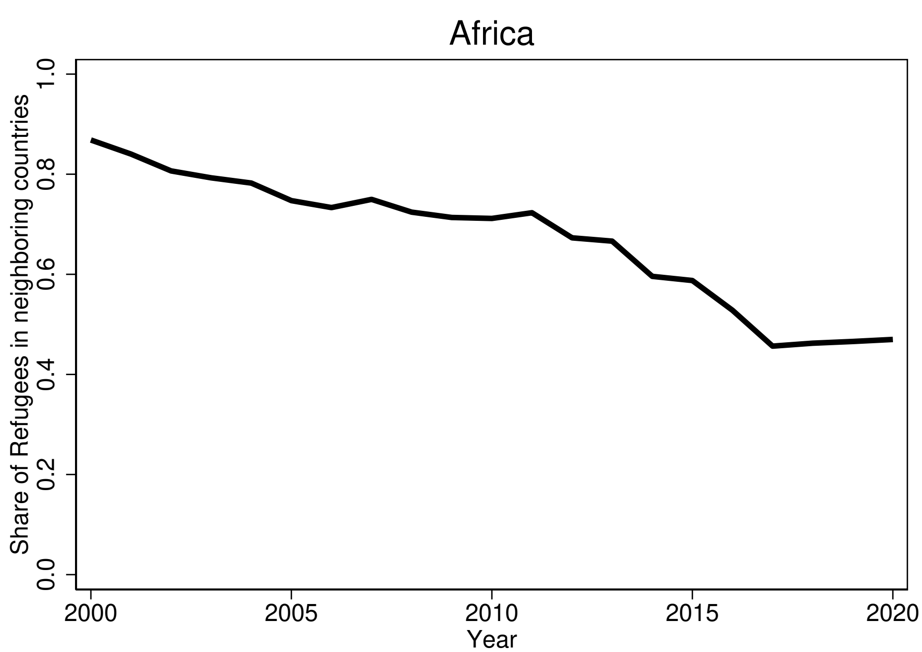
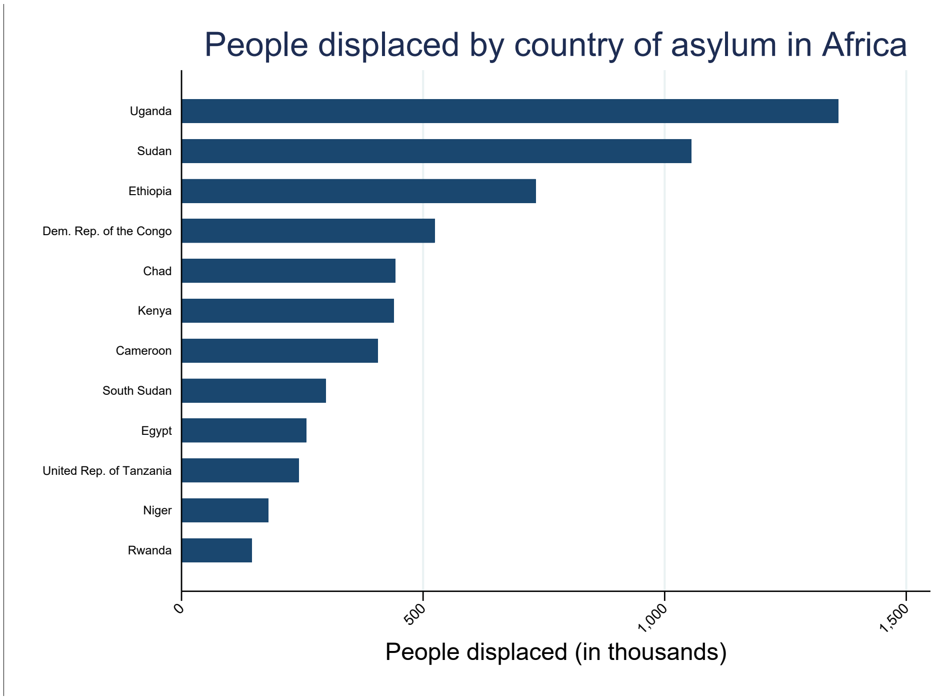
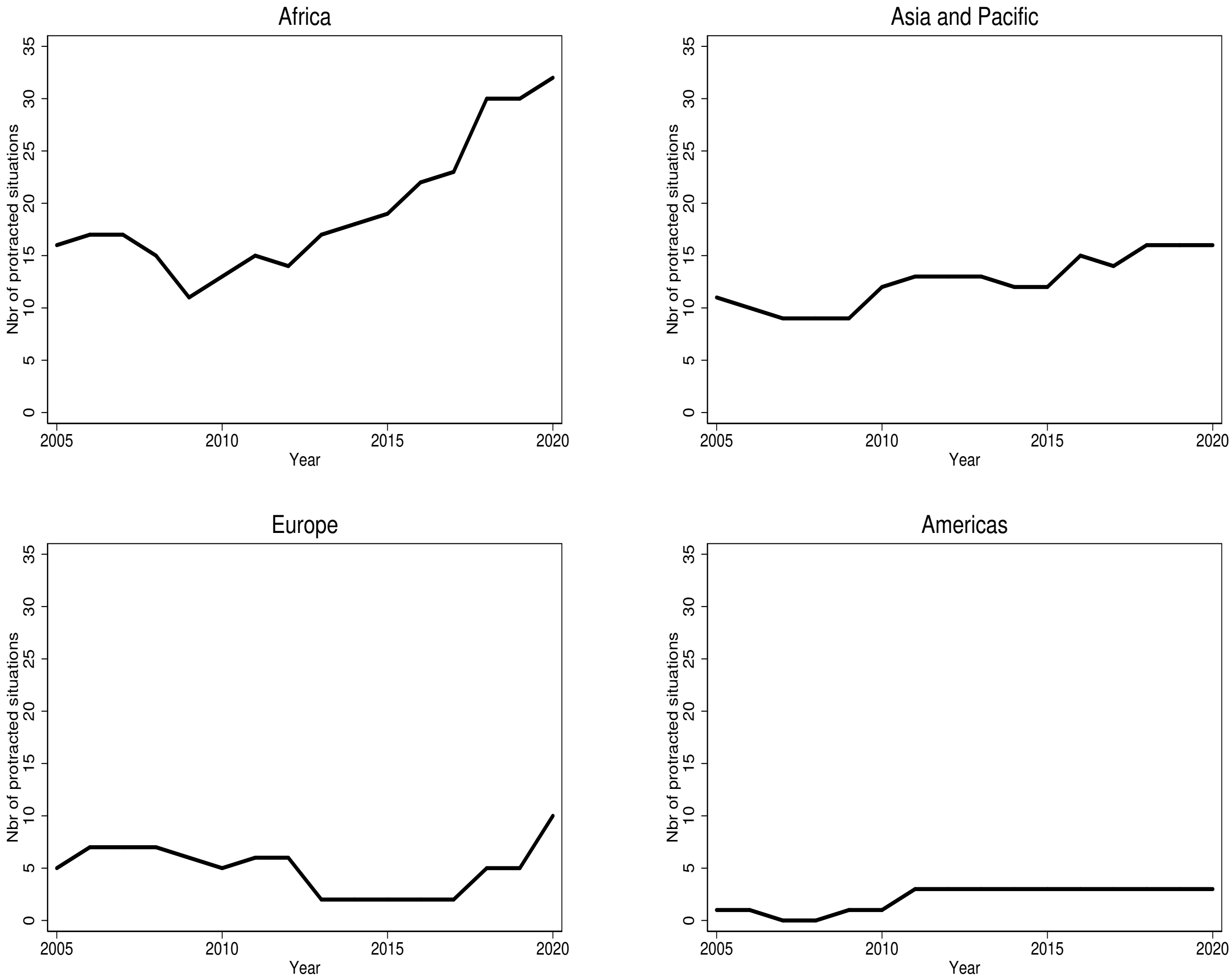


Figure D.9: People Displaced Across Borders by Country of Asylum, UNHCR, End of 2019



Note: Countries hosting more than 100,000 refugees.

Figure D.10: UNHCR Number of Protracted Refugee Situations



Data are aggregated directly from UNHCR camp-level data.

Figure D.11: Injective relations

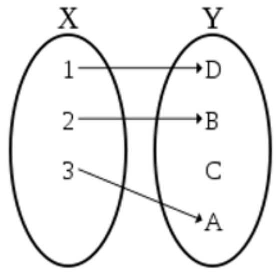


Figure D.12: Bijective relations