

Supplementary Information for:
“Ethnic diversity and conflict in Sub-Saharan Africa:”*

Luisito Bertinelli[†] Rana Cömertpay[‡] Jean-François Maystadt[§]

October 15, 2024

Abstract

This document contains a set of appendices with supplemental material.

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.

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

¹While 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

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

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

¹These 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.

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

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 LEDA² 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 ≤ 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:

$$\sum \widehat{Ref}_{cgt} = Ref_{odct} \widehat{Share}_{odgt}, \quad (C.2)$$

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

¹In 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.

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

³The 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:

$$REF_{jt} = \alpha_j + \tau_t + \delta_1 \widehat{EF}_{jt} + \delta_2 \widehat{EP}_{jt} + \delta_3 Refuges_{jt} + \delta_5 Q_{jt} + \epsilon_{1,jt} \quad (C.3)$$

and

$$REP_{jt} = \alpha_j + \tau_t + \delta_1 \widehat{EF}_{jt} + \delta_2 \widehat{EP}_{jt} + \delta_3 Refuges_{jt} + \delta_5 Q_{jt} + \epsilon_{2,jt}. \quad (C.4)$$

Appendix D Tables and Figures

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

Data	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	The Afrobarometer							UNHCR refugee camps	EPR-ER
Period	Round 1 ^I	Round 2 ^I	Round 3	Round 4	Round 5	Round 6	Round 7 ^{II}		
	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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Obs.	Mean	Std. Dev.	Min.	Max.	Obs.	Mean	Std. Dev.	Min.	Max.
<u>Conflict Events.</u>	<u>Panel A: Refugee-Hosting Areas</u>					<u>Panel B: Non-Hosting Areas</u>				
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, 80km)	2,058	0.7203	0.9183	0	4.1591	9,060	1.0214	1.2296	0	6.5309
Civilian conflict, incidence (80km)	2,327	0.5239	0.4995	0	1	12,098	0.6456	0.4784	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
<u>Diversity Indices</u>										
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
<u>Other variables</u>										
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** (0.1849)	-0.3284* (0.1859)	-0.3051 (0.1866)	-0.3208* (0.1867)
Residualized REP (Min. Ling. Dist., 80km)	1.6482*** (0.4994)	1.5492*** (0.4996)	1.5237*** (0.4999)	1.5434*** (0.4998)
Treatment group × Residualized REF	-0.0823 (0.4866)	0.1238 (0.4886)	-0.0247 (0.4994)	-0.0481 (0.4992)
Treatment group × Residualized REP	-1.7524 (1.2737)	-1.8830 (1.2732)	-1.7607 (1.2760)	-1.7013 (1.2755)
Residualized Refugees (80km, IHS)		-0.0180*** (0.0041)	-0.0226*** (0.0052)	-0.0227*** (0.0052)
Treatment group × Residualized Refugees			0.0122 (0.0085)	0.0128 (0.0085)
Treatment Group	-0.0142 (0.0149)	0.0115 (0.0160)	0.0028 (0.0171)	0.0021 (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 (0.2686)	-0.0717 (0.0536)	-0.8469 (0.5522)	-0.3959* (0.2404)	-0.3019* (0.1755)	-0.2659* (0.1521)
Refugee EP (Min. Ling. Dist., 80km)	1.2526* (0.6919)	0.2488* (0.1410)	2.5450* (1.3708)	1.2417** (0.6099)	0.9252** (0.4513)	0.7927** (0.3934)
Refugees (80km)	-0.0183*** (0.0062)	-0.0509* (0.0271)	-0.0017*** (0.0006)	-0.0043** (0.0020)	-0.0059 (0.0038)	-0.0067 (0.0056)
Rain anomalies (80km)	0.0016** (0.0008)	-0.0001 (0.0002)	0.0029*** (0.0010)	0.0009 (0.0006)	0.0002 (0.0005)	-0.0001 (0.0005)
Temp anomalies (80km)	-0.0971** (0.0450)	-0.0356*** (0.0105)	0.0130 (0.0626)	-0.0686* (0.0361)	-0.0822*** (0.0301)	-0.0862*** (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 (0.1186)	-0.1251 (0.1189)				
Native EP	0.6103* (0.3178)	0.6126* (0.3186)				
Refugees (80km, IHS)		0.0004 (0.0028)		0.0007 (0.0029)		0.0003 (0.0029)
REF (Min. Ling. Dist., 80km)			-0.1747 (0.1113)	-0.1803 (0.1141)	-0.1932* (0.1115)	-0.1958* (0.1145)
REP (Min. Ling. Dist., 80km)			0.4520 (0.2926)	0.4573 (0.2939)	0.4941* (0.2938)	0.4965* (0.2953)
Rain anomalies (80km)					-0.0008** (0.0004)	-0.0008** (0.0004)
Temp anomalies (80km)					-0.0875*** (0.0230)	-0.0874*** (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 (0.2909)	0.0790 (0.2902)				
Native EP	0.0358 (0.7807)	-0.0580 (0.7809)				
Refugees (80km, IHS)		-0.0152*** (0.0057)		-0.0149** (0.0058)		-0.0148** (0.0058)
REF (Min. Ling. Dist., 80km)			-0.3133 (0.2787)	-0.1968 (0.2773)	-0.2932 (0.2770)	-0.1785 (0.2757)
REP (Min. Ling. Dist., 80km)			0.7261 (0.7289)	0.6150 (0.7210)	0.6888 (0.7240)	0.5807 (0.7164)
Rain anomalies (80km)					-0.0006 (0.0006)	-0.0007 (0.0006)
Temp anomalies (80km)					0.0997** (0.0446)	0.0955** (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	-0.0555 (0.2502)	-0.0206 (0.2520)				
Native EP	0.9755 (0.6493)	0.8740 (0.6564)				
Refugees (80km, IHS)		-0.0164*** (0.0051)		-0.0177*** (0.0053)		-0.0178*** (0.0053)
REF (Min. Ling. Dist., 80km)			-0.4675* (0.2396)	-0.3290 (0.2435)	-0.4661* (0.2392)	-0.3279 (0.2432)
REP (Min. Ling. Dist., 80km)			1.6414*** (0.5999)	1.5094** (0.6027)	1.6414*** (0.5989)	1.5113** (0.6019)
Rain anomalies (80km)					-0.0005 (0.0007)	-0.0006 (0.0007)
Temp anomalies (80km)					0.0084 (0.0391)	0.0033 (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: 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 (0.2502)	-0.0206 (0.2520)				
Native EP	0.9755 (0.6493)	0.8740 (0.6564)				
Refugees (80km, IHS)		-0.0164*** (0.0051)		-0.0177*** (0.0053)		-0.0178*** (0.0053)
REF (Min. Ling. Dist., 80km)			-0.4675* (0.2396)	-0.3290 (0.2435)	-0.4661* (0.2392)	-0.3279 (0.2432)
REP (Min. Ling. Dist., 80km)			1.6414*** (0.5999)	1.5094** (0.6027)	1.6414*** (0.5989)	1.5113** (0.6019)
Rain anomalies (80km)					-0.0005 (0.0007)	-0.0006 (0.0007)
Temp anomalies (80km)					0.0084 (0.0391)	0.0033 (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 (0.1117)	0.0500 (0.1127)				
Native EP	0.2539 (0.3000)	0.2243 (0.3035)				
Refugees (80km, IHS)		-0.0048* (0.0027)		-0.0053* (0.0027)		-0.0059** (0.0027)
REF (Min. Ling. Dist., 80km)			-0.1227 (0.1083)	-0.0817 (0.1100)	-0.1337 (0.1087)	-0.0882 (0.1105)
REP (Min. Ling. Dist., 80km)			0.4657* (0.2812)	0.4265 (0.2837)	0.5021* (0.2820)	0.4592 (0.2849)
Rain anomalies (80km)					-0.0025*** (0.0004)	-0.0026*** (0.0004)
Temp anomalies (80km)					-0.0454** (0.0216)	-0.0471** (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 (0.2829)	0.1510 (0.2823)				
Native EP	-0.0391 (0.7529)	-0.1226 (0.7531)				
Refugees (80km, IHS)		-0.0135*** (0.0051)		-0.0138*** (0.0052)		-0.0132** (0.0052)
REF (Min. Ling. Dist., 80km)			-0.2443 (0.2709)	-0.1365 (0.2696)	-0.2229 (0.2713)	-0.1205 (0.2702)
REP (Min. Ling. Dist., 80km)			0.7022 (0.7067)	0.5993 (0.6999)	0.6478 (0.7071)	0.5514 (0.7006)
Rain anomalies (80km)					0.0020*** (0.0007)	0.0019*** (0.0007)
Temp anomalies (80km)					0.0983** (0.0451)	0.0946** (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 (0.1148)	-0.0205 (0.1152)				
Native EP	0.2870 (0.3162)	0.3162 (0.3160)				
Refugees (80km, IHS)		0.0047* (0.0025)		0.0043 (0.0026)		0.0042 (0.0026)
REF (Min. Ling. Dist., 80km)			-0.0980 (0.1084)	-0.1313 (0.1117)	-0.0978 (0.1085)	-0.1306 (0.1118)
REP (Min. Ling. Dist., 80km)			0.5567* (0.2990)	0.5884* (0.3022)	0.5570* (0.2992)	0.5879* (0.3024)
Rain anomalies (80km)					-0.0002 (0.0004)	-0.0001 (0.0004)
Temp anomalies (80km)					0.0018 (0.0211)	0.0030 (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 (0.0422)	-0.0492 (0.0410)				
Native EP	0.1572 (0.1099)	0.1352 (0.1073)				
Refugees (80km, IHS)		-0.0031* (0.0017)		-0.0039** (0.0016)		-0.0039** (0.0016)
REF (Min. Ling. Dist., 80km)			-0.0603 (0.0519)	-0.0231 (0.0479)	-0.0632 (0.0522)	-0.0260 (0.0481)
REP (Min. Ling. Dist., 80km)			0.3401*** (0.1313)	0.3011** (0.1285)	0.3448*** (0.1322)	0.3062** (0.1292)
Rain anomalies (80km)					0.0001 (0.0001)	0.0000 (0.0001)
Temp anomalies (80km)					-0.0126 (0.0082)	-0.0143* (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 (0.1007)	-0.0582 (0.1020)				
Native EP	0.2537 (0.2829)	0.1883 (0.2835)				
Refugees (80km, IHS)		-0.0106*** (0.0023)		-0.0114*** (0.0023)		-0.0114*** (0.0023)
REF (Min. Ling. Dist., 80km)			-0.0880 (0.0973)	0.0012 (0.1021)	-0.0799 (0.0978)	0.0080 (0.1027)
REP (Min. Ling. Dist., 80km)			0.3353 (0.2661)	0.2502 (0.2741)	0.3191 (0.2673)	0.2363 (0.2754)
Rain anomalies (80km)					-0.0000 (0.0003)	-0.0001 (0.0003)
Temp anomalies (80km)					0.0395* (0.0214)	0.0362* (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 (0.2270)	-0.0885 (0.2278)				
Native EP	0.0461 (0.5764)	0.0980 (0.5792)				
Refugees (80km, IHS)		0.0084** (0.0038)		0.0092** (0.0041)		0.0095** (0.0041)
REF (Min. Ling. Dist., 80km)			-0.0631 (0.2035)	-0.1352 (0.2119)	-0.0482 (0.2023)	-0.1219 (0.2108)
REP (Min. Ling. Dist., 80km)			0.1188 (0.5062)	0.1876 (0.5119)	0.0865 (0.5033)	0.1559 (0.5089)
Rain anomalies (80km)					0.0004 (0.0005)	0.0004 (0.0005)
Temp anomalies (80km)					0.0714** (0.0282)	0.0741*** (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 (0.0823)	0.0444 (0.0827)				
Native EP	-0.1593 (0.2143)	-0.1351 (0.2154)				
Refugees (80km, IHS)		0.0039*** (0.0013)		0.0040*** (0.0014)		0.0041*** (0.0014)
REF (Min. Ling. Dist., 80km)			0.0473 (0.0747)	0.0158 (0.0775)	0.0574 (0.0742)	0.0255 (0.0769)
REP (Min. Ling. Dist., 80km)			-0.1039 (0.1908)	-0.0739 (0.1923)	-0.1229 (0.1891)	-0.0928 (0.1903)
Rain anomalies (80km)					-0.0002 (0.0003)	-0.0002 (0.0003)
Temp anomalies (80km)					0.0498*** (0.0152)	0.0509*** (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 (0.1954)	0.1215 (0.1959)				
EP (no dist.)	-0.1074 (0.5490)	-0.1829 (0.5502)				
Refugees (80km, IHS)		-0.0180*** (0.0058)		-0.0170*** (0.0065)		-0.0171*** (0.0065)
REF (Min. Ling. Dist., 80km) (no dist.)			-0.4397** (0.1980)	-0.3029 (0.2072)	-0.4470** (0.1995)	-0.3105 (0.2087)
REP (Min. Ling. Dist., 80km) (no dist.)			0.9780* (0.5134)	0.8947* (0.5152)	1.0042* (0.5171)	0.9221* (0.5186)
Rain anomalies (80km)					0.0017** (0.0008)	0.0016** (0.0008)
Temp anomalies (80km)					-0.0931** (0.0451)	-0.0981** (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 (0.1256)	-0.1008 (0.1255)				
EP	1.0532** (0.4515)	1.0022** (0.4516)				
Refugees (80km, IHS)		-0.0175*** (0.0059)		-0.0164** (0.0065)		-0.0165** (0.0065)
REF (Min. Ling. Dist., 80km) (no dist.)			-0.3065*** (0.1177)	-0.1807 (0.1263)	-0.2998** (0.1182)	-0.1735 (0.1270)
REP (Min. Ling. Dist., 80km)			1.0473** (0.4150)	0.9337** (0.4162)	1.0166** (0.4171)	0.9029** (0.4184)
Rain anomalies (80km)					0.0016** (0.0008)	0.0016** (0.0008)
Temp anomalies (80km)					-0.0894** (0.0450)	-0.0946** (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 (0.3156)	-0.0148 (0.3190)				
EP	1.1331 (0.7453)	1.0292 (0.7518)				
EF (no dist.)	-0.1039 (0.1440)	-0.0972 (0.1442)				
Refugees (80km, IHS)		-0.0175*** (0.0059)		-0.0166** (0.0064)		-0.0167*** (0.0064)
Refugee EF (Min. Ling. Dist., 80km) (no dist.)			-0.2890** (0.1406)	-0.1480 (0.1475)	-0.2713* (0.1408)	-0.1289 (0.1479)
Refugee EF (Min. Ling. Dist., 80km)			-0.0640 (0.3098)	-0.1161 (0.3111)	-0.1042 (0.3116)	-0.1579 (0.3127)
Refugee EP (Min. Ling. Dist., 80km)			1.1579* (0.6817)	1.1332* (0.6856)	1.1964* (0.6877)	1.1741* (0.6911)
Rain anomalies (80km)					0.0016** (0.0008)	0.0016** (0.0008)
Temp anomalies (80km)					-0.0903** (0.0450)	-0.0959** (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 (0.2195)	-0.1650 (0.2213)				
EP	0.7634 (0.5990)	0.7246 (0.6030)				
Refugees (40km, IHS)		-0.0191*** (0.0059)		-0.0190*** (0.0060)		-0.0188*** (0.0060)
REF (Min. Ling. Dist., 40km)			-0.3903* (0.2168)	-0.3011 (0.2189)	-0.3875* (0.2167)	-0.2999 (0.2188)
REP (Min. Ling. Dist., 40km)			1.1855** (0.5801)	1.0758* (0.5826)	1.1711** (0.5799)	1.0642* (0.5823)
Rain anomalies (40km)					0.0013** (0.0006)	0.0012** (0.0006)
Temp anomalies (40km)					0.0087 (0.0375)	0.0066 (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 (0.3034)	-0.2867 (0.3039)				
EP	1.4510* (0.8206)	1.4406* (0.8216)				
Refugees (120km, IHS)		-0.0025 (0.0052)		-0.0020 (0.0055)		-0.0024 (0.0055)
REF (Min. Ling. Dist., 120km)			-0.4693* (0.2701)	-0.4516 (0.2769)	-0.5079* (0.2731)	-0.4867* (0.2801)
REP (Min. Ling. Dist., 120km)			1.3256* (0.7026)	1.3151* (0.7041)	1.3962** (0.7112)	1.3839* (0.7127)
Rain anomalies (120km)					0.0012 (0.0009)	0.0012 (0.0009)
Temp anomalies (120km)					-0.1828*** (0.0492)	-0.1838*** (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	-0.0379 (0.1944)	-0.0114 (0.1649)				
EP	0.7190 (0.5031)	0.6514 (0.4476)				
Refugees (80km, IHS)		-0.0189*** (0.0038)		-0.0208*** (0.0047)		-0.0188*** (0.0045)
REF (Min. Ling. Dist., 80km)			-0.1630 (0.1815)	-0.0522 (0.1845)	-0.2019 (0.1822)	-0.1327 (0.1878)
REP (Min. Ling. Dist., 80km)			0.7003 (0.4604)	0.6549 (0.4609)	0.7689* (0.4625)	0.7087 (0.4728)
Rain anomalies (80km)					-0.0004 (0.0006)	-0.0002 (0.0006)
Temp anomalies (80km)					-0.1990*** (0.0255)	-0.0601* (0.0353)
Observations	23,236	23,236	23,236	23,236	23,236	23,236
R-squared	0.779	0.779	0.778	0.779	0.779	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 (0.1333)	-0.1250 (0.1319)				
EP	0.6276* (0.3497)	0.6351* (0.3458)				
Refugees (80km, IHS)		0.0014 (0.0018)		0.0021 (0.0020)		0.0017 (0.0022)
REF (Min. Ling. Dist., 80km)			-0.1840 (0.1271)	-0.2184* (0.1311)	-0.2215 (0.1402)	-0.2478* (0.1441)
REP (Min. Ling. Dist., 80km)			0.4973 (0.3282)	0.5395 (0.3282)	0.5843 (0.3613)	0.6157* (0.3610)
Rain anomalies (80km)					-0.0008** (0.0004)	-0.0008* (0.0004)
Temp anomalies (80km)					-0.0818** (0.0338)	-0.0794** (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 (0.2750)	-0.1441 (0.2775)				
EP	3.4418*** (0.9132)	2.7162*** (0.9705)				
Refugees (80km, IHS)		-0.0139** (0.0064)		-0.0198*** (0.0061)		-0.0195*** (0.0061)
REF (Min. Ling. Dist., 80km)			-0.4186 (0.2618)	-0.2693 (0.2684)	-0.4340* (0.2635)	-0.2874 (0.2700)
REP (Min. Ling. Dist., 80km)			2.3175** (0.9267)	2.5794*** (0.9823)	2.2450** (0.9363)	2.5072** (0.9912)
Rain anomalies (80km)					0.0018** (0.0008)	0.0016** (0.0008)
Temp anomalies (80km)					-0.1076** (0.0452)	-0.1117** (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 (0.2795)	-0.1288 (0.2823)				
EP	1.2241 (0.7596)	1.1172 (0.7662)				
Refugees (80km, IHS)		-0.0183*** (0.0059)		-0.0192*** (0.0062)		-0.0192*** (0.0062)
REF (Min. Ling. Dist., 80km)			-0.4357* (0.2645)	-0.2864 (0.2713)	-0.4514* (0.2661)	-0.3036 (0.2728)
REP (Min. Ling. Dist., 80km)			1.3515* (0.6970)	1.2096* (0.7016)	1.3743* (0.7022)	1.2360* (0.7062)
Rain anomalies (80km)					0.0017** (0.0008)	0.0016** (0.0008)
Temp anomalies (80km)					-0.0952** (0.0465)	-0.1011** (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	0.9463***	0.9536***	0.9456***	0.9542***
	(0.0086)	(0.0101)	(0.0084)	(0.0094)
Predicted REP	0.0501**	0.0426*	0.0447**	0.0367*
	(0.0218)	(0.0227)	(0.0211)	(0.0212)
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*** (0.0003)	0.0008*** (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 (0.0005)	-0.0014** (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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Obs.	Mean	Std. Dev.	Min.	Max.	Obs.	Mean	Std. Dev.	Min.	Max.
	Panel A: Refugee-Hosting Areas					Panel B: All Areas				
<u>Diversity Indices</u>										
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
<u>Sociodemographic Variables</u>										
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
<u>Outcome Variables</u>										
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
<u>Climate Data</u>										
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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Attack ^a	Crime	Gen. Trust	Gov. Trust	National Id.	Neigh. Trust	Protest	Theft
REF (80 km, min. ling. dist.)	-0.0762 (0.0524)	-0.1497 (0.1013)	-0.1384 (0.1958)	0.0714 (0.0921)	-0.0662 (0.1220)	-0.2880 (0.1914)	-0.1157* (0.0694)	-0.1149 (0.0877)
REP (80 km, min. ling. dist.)	0.3162** (0.1584)	0.3142 (0.2789)	0.1978 (0.5207)	0.0228 (0.2510)	0.1069 (0.3358)	0.4818 (0.5988)	0.2181 (0.2011)	0.1169 (0.2493)
Refugees (80 km, IHS)	0.0009 (0.0015)	0.0020 (0.0042)	-0.0030 (0.0071)	0.0035 (0.0033)	0.0035 (0.0045)	-0.0005 (0.0110)	-0.0027 (0.0033)	0.0012 (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 (0.0550)	-0.0588 (0.0549)				
EP	0.2636 (0.1615)	0.2735* (0.1613)				
Refugees (80 km, IHS)		0.0013 (0.0014)		0.0013 (0.0015)		0.0009 (0.0015)
REF (80 km, min. ling. dist.)			-0.0630 (0.0525)	-0.0706 (0.0533)	-0.0713 (0.0514)	-0.0762 (0.0524)
REP (80 km, min. ling. dist.)			0.2927* (0.1587)	0.3059* (0.1582)	0.3078* (0.1587)	0.3162** (0.1584)
Rain anomalies					0.0005 (0.0004)	0.0005 (0.0004)
Temp. anomalies					-0.0388 (0.0246)	-0.0378 (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

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

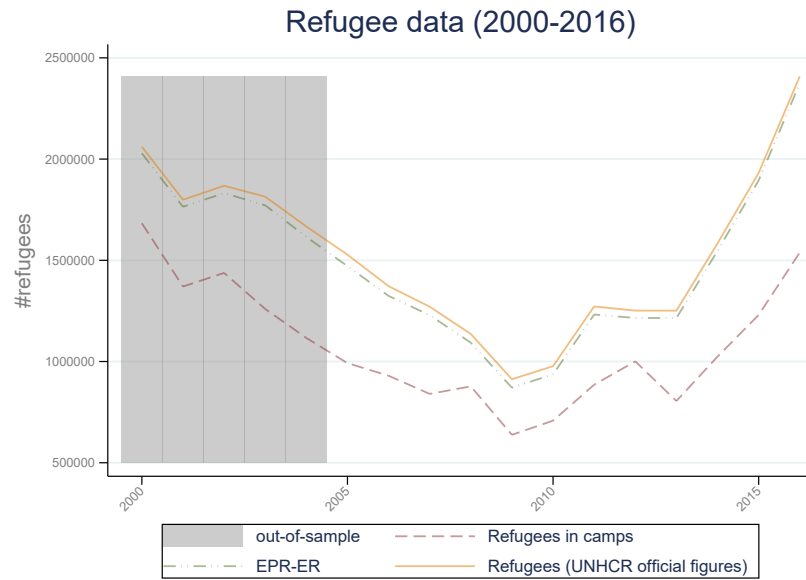


Figure D.2: Refugee Flows between Source and Asylum Countries

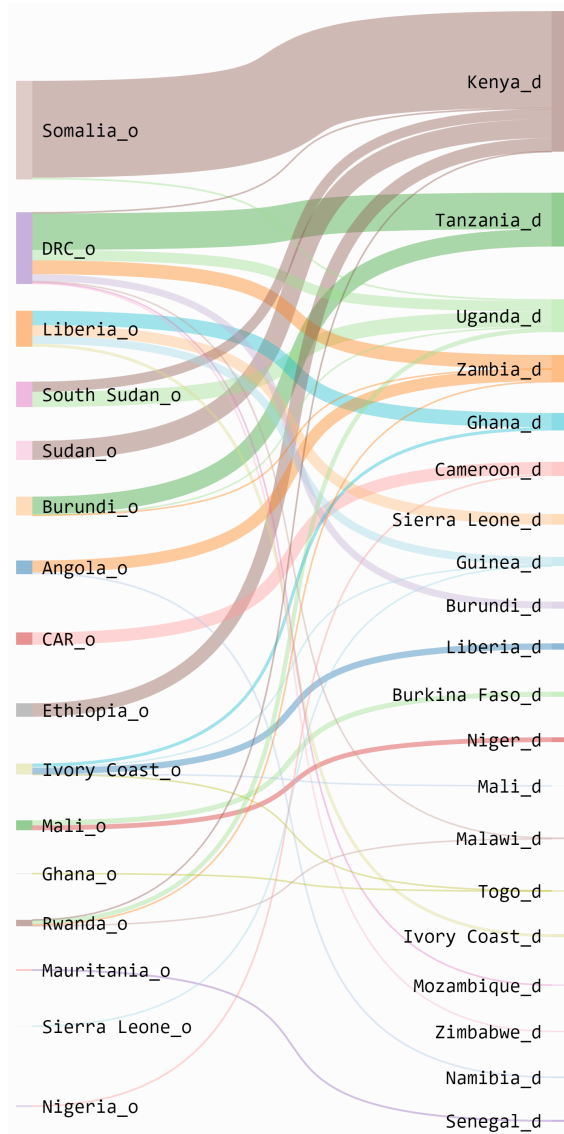
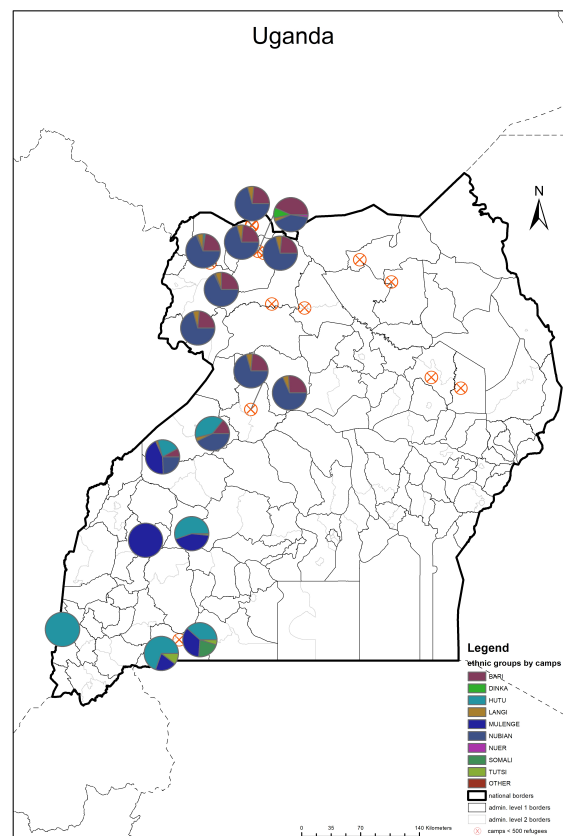
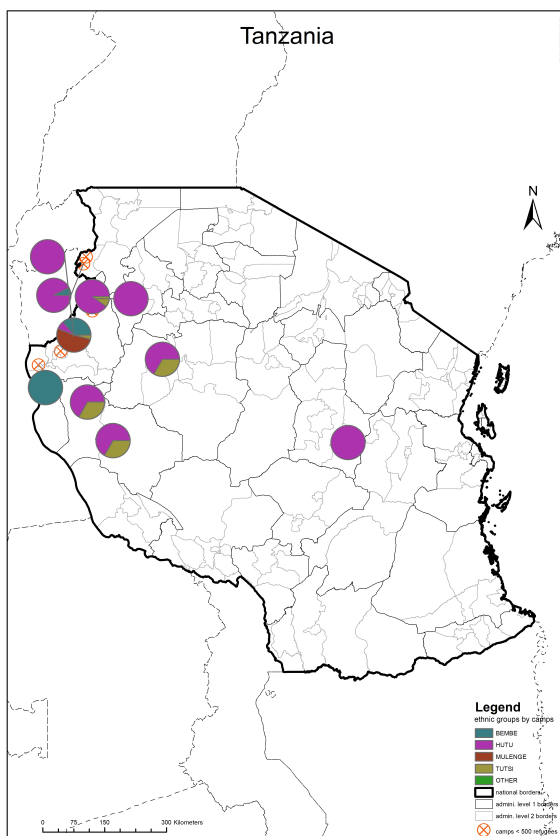
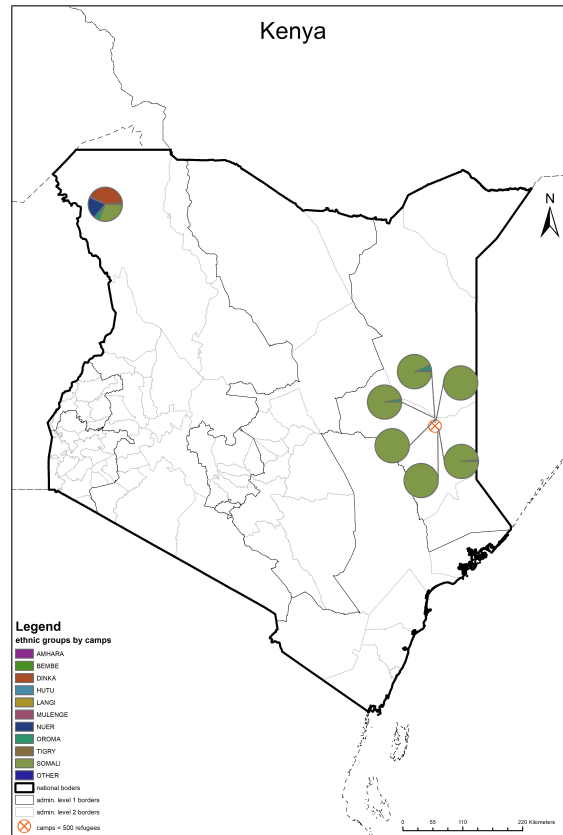
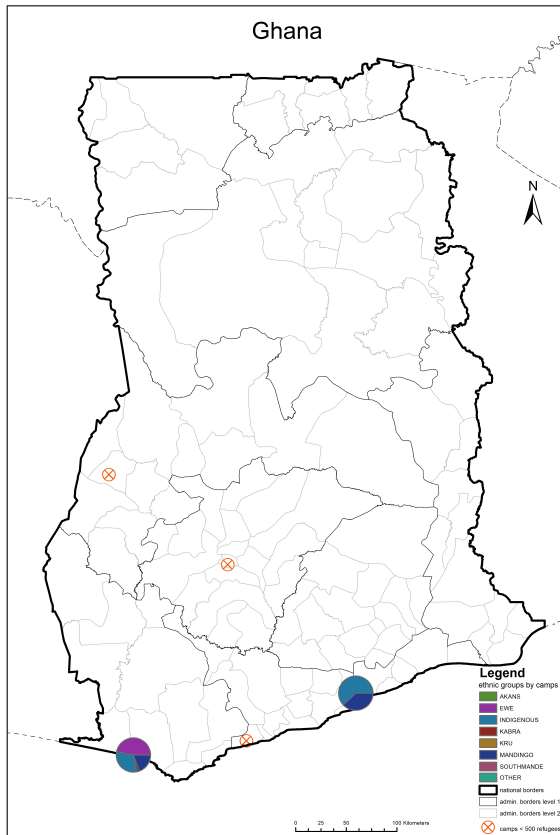


Figure D.3: Refugees in Camps per Ethnic Group



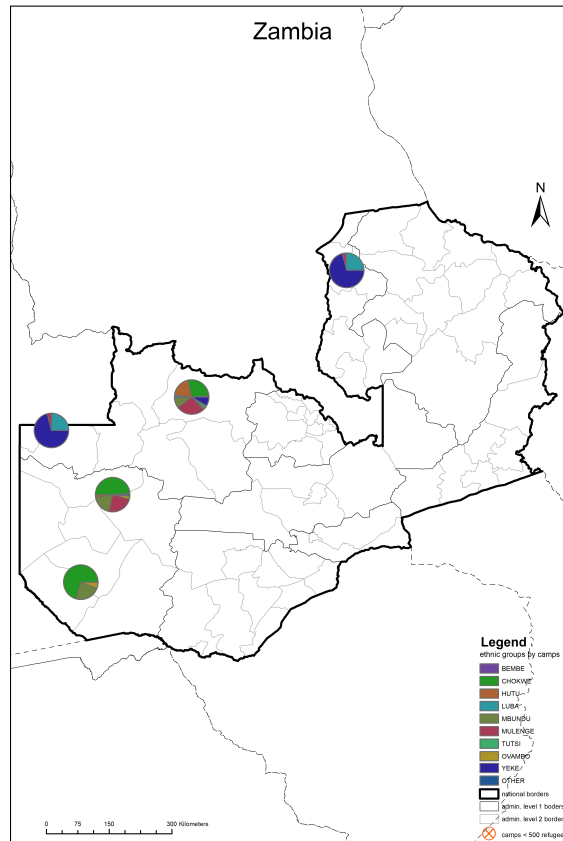
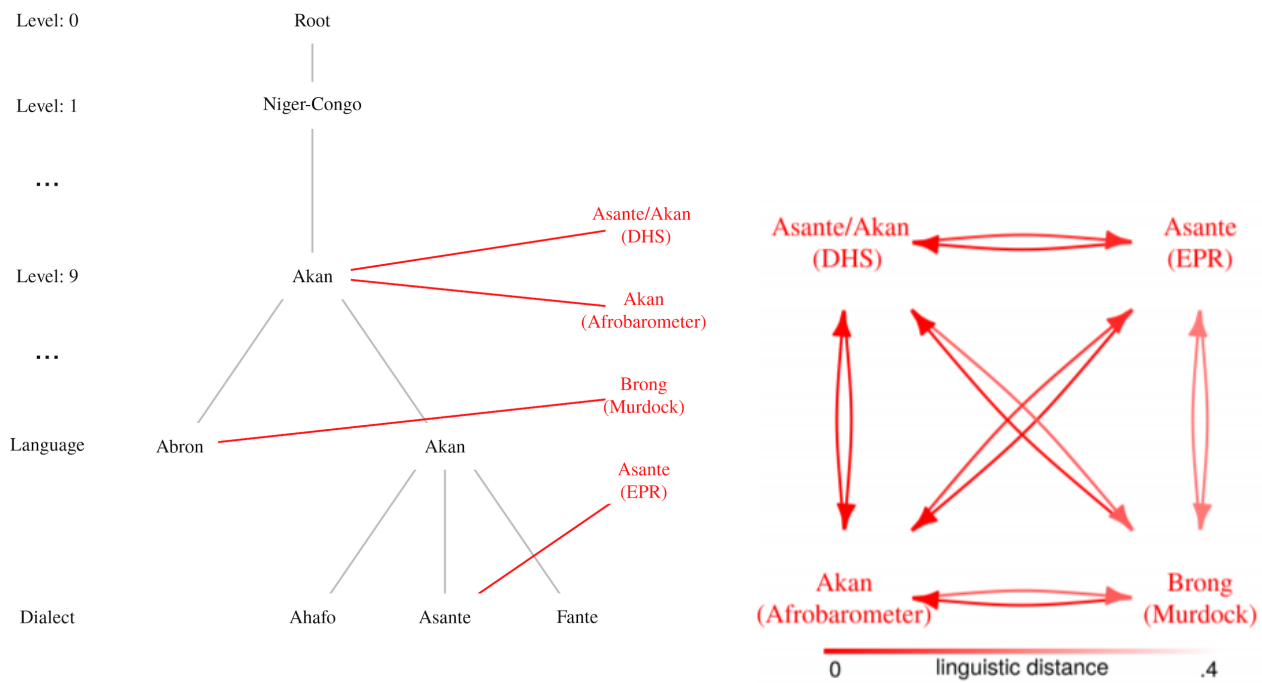


Figure D.4: Linking Ethnic Data from Africa



Source: Müller-Crepon et al., 2020.

Figure D.5: Heterogeneity Analysis

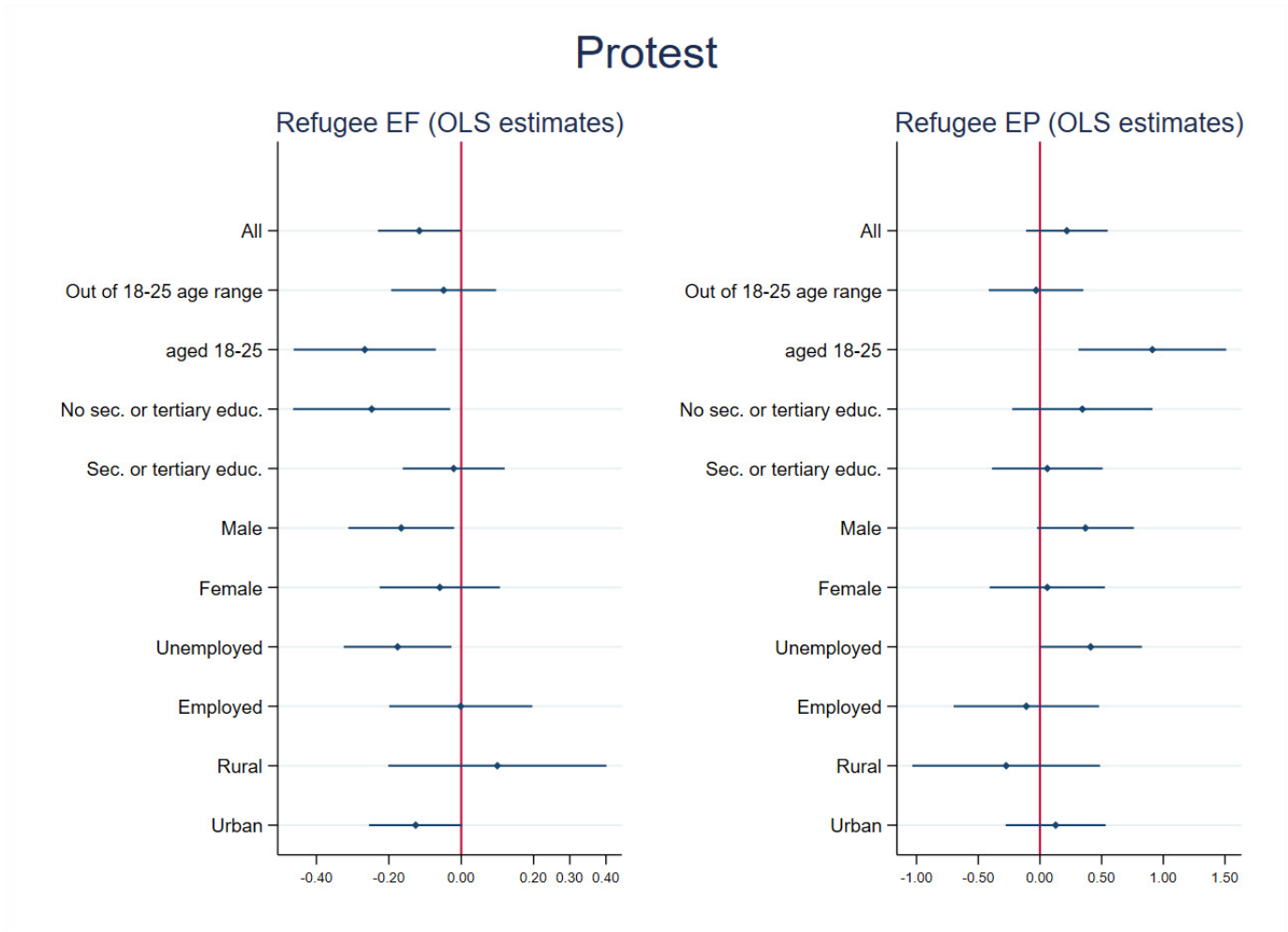


Figure D.6: UNHCR Aggregated Refugee Data by Region of Asylum

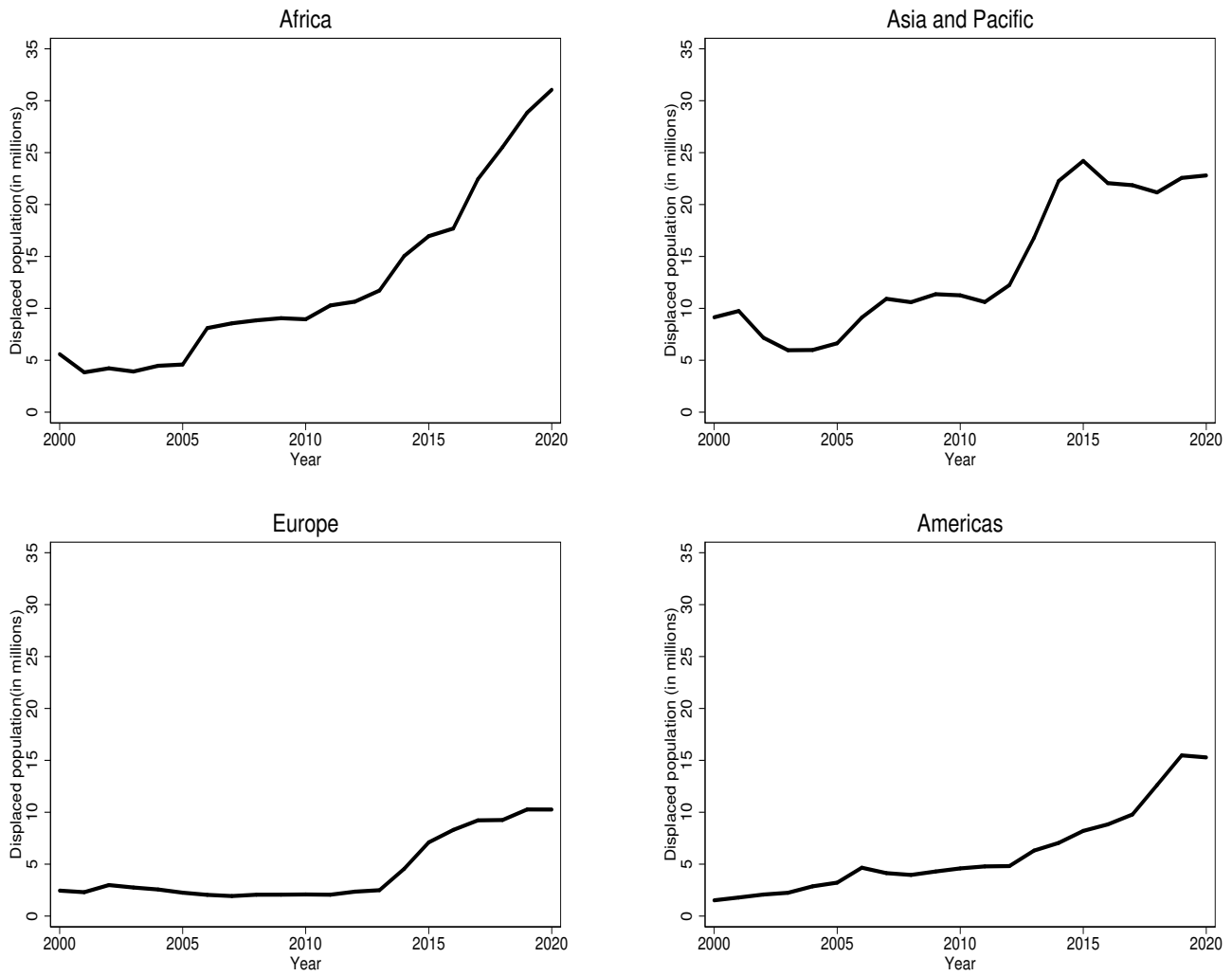
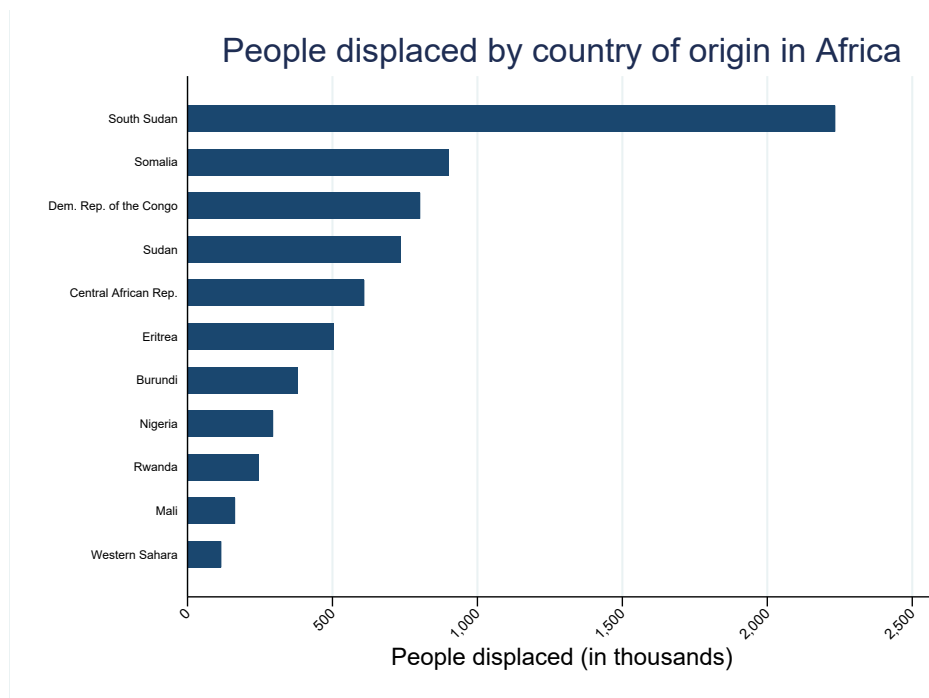


Figure D.7: People Displaced Across Borders by Country of Origin, UNHCR, End of 2019



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

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

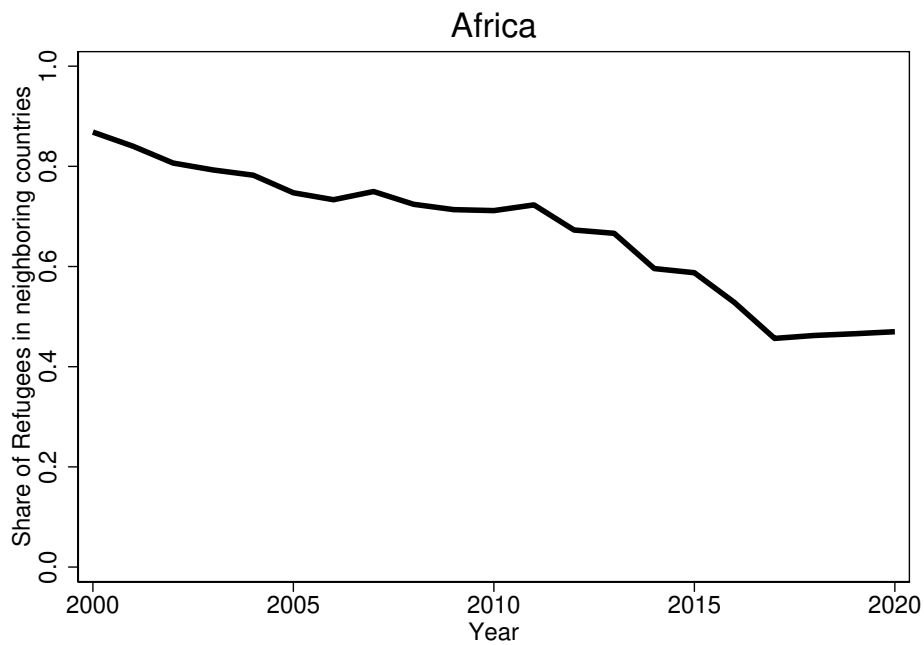
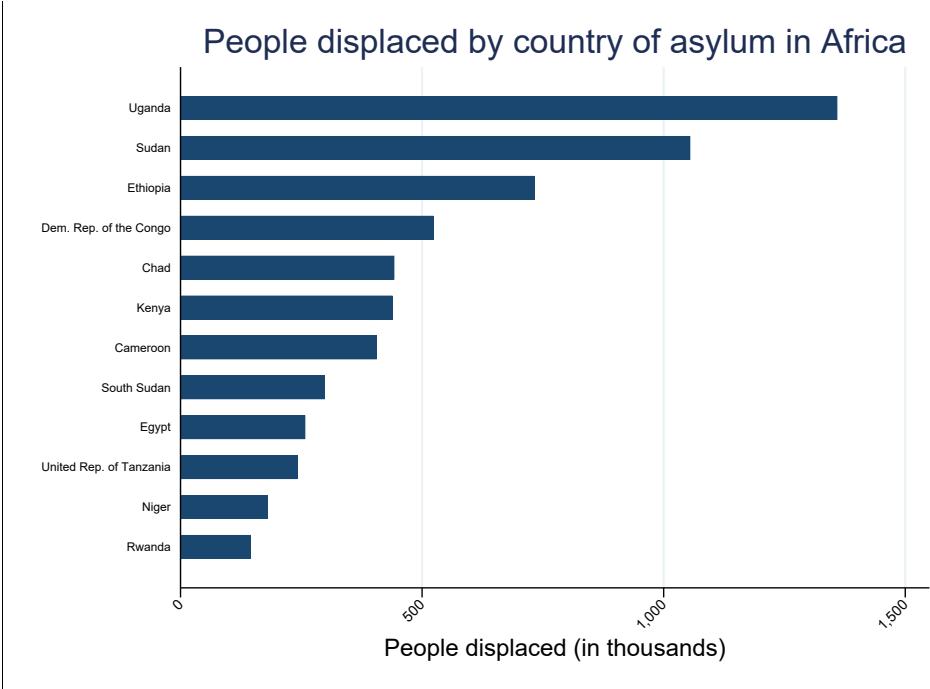
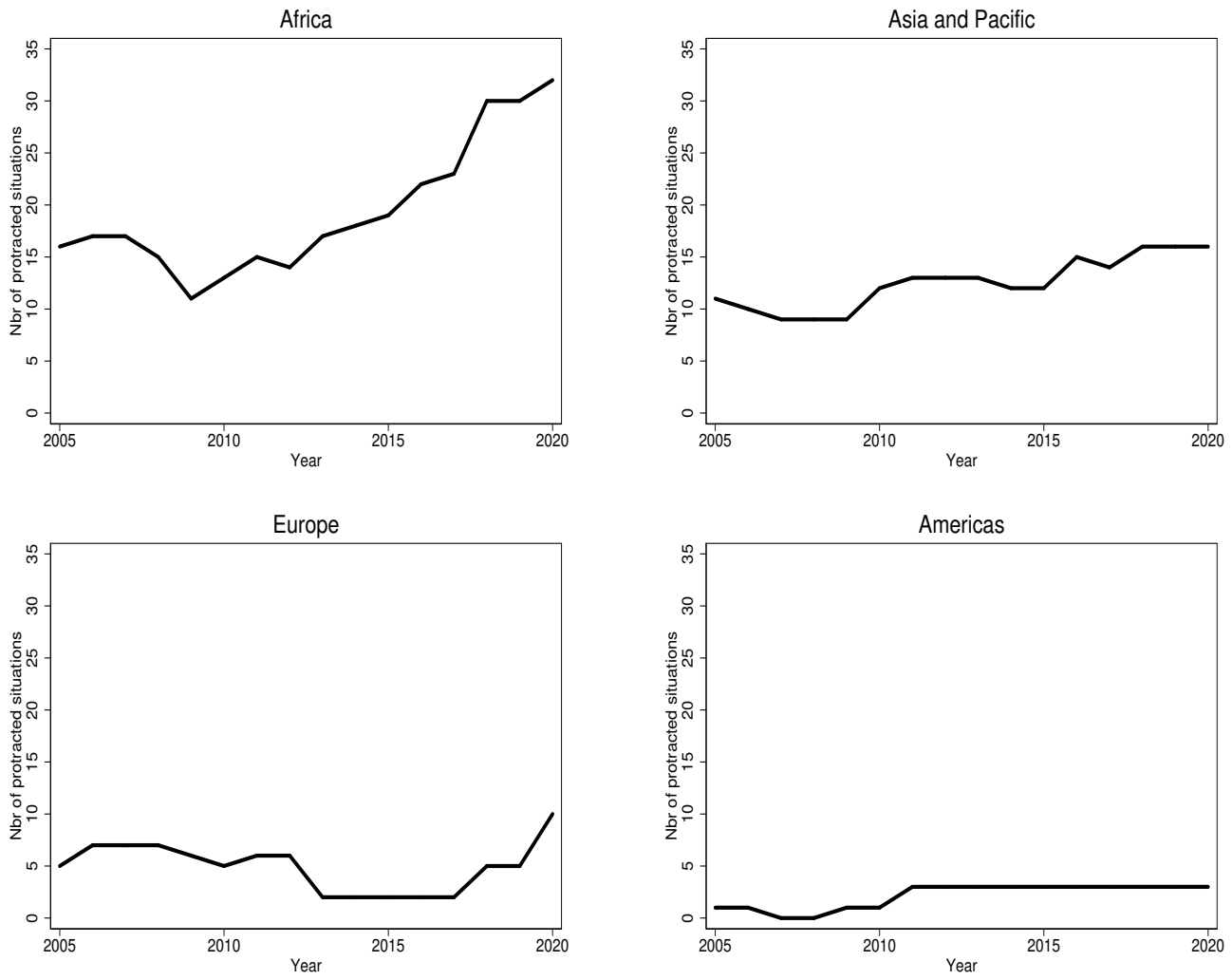


Figure D.9: People Displaced Across Borders by Country of Asylum, UNHCR, End of 2019



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

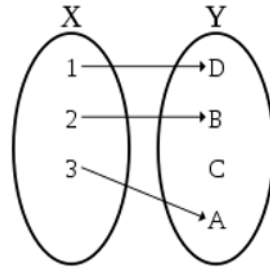
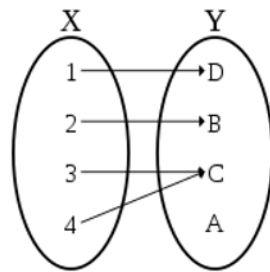


Figure D.12: Bijective relations



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>.
- 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.
- 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.
- Crozet, M. (2004). Do Migrants Follow Market Potentials? An Estimation of a New Economic Geography Model. Journal of Economic Geography 4(4), 439–458.
- Devictor, X., Q.-T. Do, and A. Levchenko (2021). The globalization of refugee flows. Journal of Development Economics 150, 102605.
- 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.
- Mayda, A. M. (2010). International migration: A panel data analysis of the determinants of bilateral flows. Journal of Population Economics 23(4), 1249–1274.
- Murdock, G. (1967). The ethnographic atlas: a summary. Ethnology 6, 109–236.
- 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.
- United Nations High Commissioner for Refugees (2020). Global Trends: Forced Displacement in 2019. Geneva.
- Verwimp, P. and J. Maystadt (2015). Forced Displacement and Refugees in Sub-Saharan Africa. An Economic Inquiry. World Bank Policy Research Working Paper 7517.
- Zylkin, T. (2019). ppml with negative values. Statalist.