THE GRAVITY OF VIOLENCE*

– ONLINE APPENDIX –

Mathieu Couttenier† Julian Marcoux‡ Thierry Mayer§ Mathias Thoenig¶

September 5, 2024

# OA1 The Armed Conflict Location & Event Data Project (ACLED)

Our aim is to process geolocalized violent events into bilateral flows of violence with well-defined origins and destinations. We detail our data construction procedure, which consists of six steps, outlined in Table OA1.1.

**General Overview** We use conflict event data from the Armed Conflict Location and Event Dataset (ACLED) which contains information on the geo-location of conflict events in all African countries over the period from 1997 to June $23^{th}$, 2023.[1] In the raw data, the unit of observation is a violent event for which we have information about the nature of the actors involved[2], the nature of the events[3], the location (longitude and latitude), fatalities and the precise date. The dataset covers 325541 distinct events spread over 57 countries. For each event, the database reports the involved

*Corresponding author: Mathias Thoenig (mathias.thoenig@unil.ch).

†Department of Economics, ENS de Lyon; and CEPR.

‡Department of Economics, University of Lausanne.

§Department of Economics, Sciences Po, Paris; CEPII; and CEPR.

¶Department of Economics, University of Lausanne; IMD; and CEPR.

[1] The dataset "*Compatibility*" has been download on June 28, 2023 (2pm).

[2] ACLED defines 8 natures of actors. "Rebel groups" are violent political entities aiming to challenge national government. "Political militias" encompass actors that for specific objectives and/or a defined period of time make use of violence to achieve some political objectives. "State forces" (military and police units) are actors executing governmental functions. "Identity Militias" are actors organized around a collective motives (e.g. community, ethnicity, region, religion, livelihood). "Rioters" are actors engage in violence either during demonstrations or in spontaneous acts of violence. "Protesters" are peaceful demonstrators. "Civilians" are victims of violence. "External/Other Forces" encompass different actors such as to international organizations, state forces active outside of their main country, private security and mercenaries

[3] They record six event types (and 25 sub-event types). "Battles" are defined as a violent interaction between two actors. "Explosions/Remote violence" correspond to incidents in which one side uses weapon types that are at widely destructive. "Protests" corresponds to non-violent meeting with more than three participants. "Riots" are violence committed by groups of three or more demonstrators or mobs engaging in destructive behavior. "Strategic developments" encompasses events involving groups that are not classified as "Political violence" or "Demonstrations", but may have the potential to influence political dynamics within and between states. "Violence against civilians" correspond to violent events where an organized armed group inflicts violence upon unarmed non-combatants

actors under the categories *actor1*, *actor2*, *associated actor1*, and *associated actor2*.[4] We refer to these four categories collectively as the actors. In the raw data, there are 14137 distinct actors.

Table OA1.1: Data-processing steps

| Steps | # Events | # Actors | # Fatalities |
|---|---|---|---|
| 0-Raw dataset | 325541 | 14137 | 882975 |
| 1- Geographic filter | 237489 | 11138 | 556016 |
| 2- Events filter | 234943 | 11013 | 555959 |
| 3- Information filter | 232878 | 5486 | 553745 |
| 4- Actors selection | 87578 | 1492 | 379314 |
| 5- Murdock filter on actors | 80619 | 220 | 362672 |
| 6- Murdock filter on events | 78335 | 220 | 353289 |

Note: **Step 0–Raw dataset:** reshaped from raw data such that the unit of observation is an event×actor(group). **Step 1–Geographic filter:** only consider events with the highest level of precision and exclude events outside continental Africa. **Step 2–Events filter:** exclude events considered as peaceful or with unidentified nature. **Step 3–Information filter:** exclude actors with no external information (e.g., no information on the nature of the actors). **Step 4–Actor selection:** keep actors identified as “Rebel groups” and “Political militias”. **Step 5- Murdock filter on actors:** 220 actors that we assign to 87 Murdock ethnic groups. **Step 6-Murdock filter on events:** exclude events that do not intersect with the Murdock map (Sinaï region, lakes...).

**Step 0-Raw dataset:** We first reshape the raw data so that the unit of observation becomes an actor×event cell. We end up with 624571 observations. Tables OA1.2 and OA1.3 display respectively at the different steps of the data-processing the percentage of event by nature of actors[5] and the percentage of event by nature of violence.

[4]For a given event, information on the nature of the actors (see footnote 2) is reported only for those assigned to the categories *actor1* and *actor2*. For those assigned to the other two categories, we recover this information by cross-matching their names with those reported in the *actor1* and *actor2* categories for *other* events, exploiting the fact that an actor can be assigned to *actor1* in one event but to *assoc_actor1* in another. However, we are not able to recover this information for all actors. Here is the list of the 10 actors with the highest number of events for which we do not have information on their nature: Labor Group (Morocco); Labor Group (Nigeria); Labor Group (South Africa); Labor Group (Tunisia); Oromo Ethnic Group (Ethiopia); Pastoralists (Nigeria); Resistance Committees (Sudan); Students (South Africa); Women (Democratic Republic of Congo)

[5]Note that the columns do not sum to 100% because some events involve actors which are associated to at least two different natures. Here is the list of actors with multiple natures: Darul-Islam (Nigeria); Dinka Ethnic Militia (South Sudan); FDPC: Democratic Front for the People of the Central African Republic; FNL: National Forces of Liberation; HSGF: Homeland Study Group Foundation; MAPI: Ituri Popular Self-Defense Movement; Mayi Mayi Militia (UPLC: Union of Patriots for the Liberation of Congo); NDA: National Democratic Alliance (Sudan); Ngumino Ethnic Militia (Democratic Republic of Congo); OLF: Oromo Liberation Front (Shane Splinter Faction); RRR: Return, Reclamation and Rehabilitation; Red Ant Security Relocation and Eviction Services; Salafist Muslim Militia (Tunisia); Sinai Tribal Union; TPLF: Tigray People’s Liberation Front; UFDR: Union of Democratic Forces for Unity; UNITA: National Union for the Total Independence of Angola.

Table OA1.2: Actors of violence

| | Step 0 | | Step 3 | |
|---|---|---|---|---|
| Nature of actors | % Events | % Actors | % Events | % Actors |
| Rebel groups | 26.5 | 2.7 | 22.9 | 5.8 |
| Political Militias | 30.8 | 10.3 | 17.4 | 21.6 |
| State Forces | 39.8 | 6.5 | 39.3 | 15.5 |
| Identity Militias | 9.8 | 29.8 | 6.4 | 50.9 |
| Rioters | 11.1 | .4 | 13.1 | 1 |
| Protesters | 23.6 | .5 | 30 | 1.1 |
| Civilians | 33.5 | .7 | 30.4 | 1.6 |
| External/Other Forces | 30.7 | 2.6 | 29.1 | 6.1 |
| Missing nature | 29.2 | 47.9 | 0 | 0 |

Note: The figures in this table are provided at different steps of the data-processing described in Table OA1.1: for the raw data (step 0) and after the information filter (step 3). For more information on the definition of the nature of the actors, see footnote 2. *Missing nature* corresponds to actors for which there is no information on the nature in the ACLED dataset.

Table OA1.3: Nature of violence

| | Step 0 | Step 4 | Step 6 |
|---|---|---|---|
| Type of violence | | % Events | |
| Battles | 25.3 | 44.8 | 49.8 |
| Explosions/Remote violence | 6.7 | 8.2 | 9.12 |
| Protests | 19.7 | 7.5 | 7.77 |
| Riots | 12.5 | 6.2 | 6.47 |
| Strategic developments | 7 | 7.6 | 8.62 |
| Violence against civilians | 28.7 | 25.6 | 29.57 |

Note: The figures in this table are provided at different steps of the data-processing described in Table OA1.1: for the raw data (step 0), after the *actors' selection* (step 4) and after the *Murdock filter on events* (step 6). For more information on the definition of the nature of violence, see footnote 3.

**Step 1- Geographic filter:** The ACLED dataset documents three levels of precision for the spatial location of events.[6] A total of 74.58% of the events are recorded with the highest level of precision (town level) while the rest are associated with lower levels of precision ( 22.95% at the small region level and 2.47% at the large region level). Table OA1.4 displays the percentage of events by actors and geo-precision. Our *geographic filter* (step 1) retains only observations with the highest level of precision (geo-precision 1). It also excludes events that are not part of continental Africa, i.e events happening at sea and on islands (such as Madagascar). This sample cut leads to 237489 distinct events covering 11138 actors.

Table OA1.4: Percentage of events by actors and geoprecision

| Nature of actors | Geo-precision 1 | Geo-precision 2 | Geo-precision 3 |
|---|---|---|---|
| Rebel groups | 63.4 | 33.2 | 3.4 |
| Political Militias | 72.7 | 24.6 | 2.7 |
| State Forces | 73.4 | 23.5 | 3.1 |
| Identity Militias | 52.3 | 45.1 | 2.5 |
| Rioters | 87 | 12.2 | .8 |
| Protesters | 93.3 | 5.6 | 1.1 |
| Civilians | 67.2 | 30.3 | 2.6 |
| External/Other Forces | 69.5 | 27 | 3.4 |
| | | | |
| Missing nature | 74.7 | 23 | 2.4 |

Note: The figures in this table are provided at the step of the data-processing described in Table OA1.1: 0-Raw dataset (step 0). The actors are defined as in the note of Table OA1.2. See footnote 6 for the definition of the levels of geo-precision.

**Step 2- Events filter:** ACLED reports information on the nature of violence attached to each event. A small fraction of events are categorized as non-violent or non-identified (sub-event: "Agreement" and sub-event: "Others", respectively). The *events filter* (step 2) excludes those events. After this filter, the sample comprises 234943 events involving 11013 actors.

**Step 3- Information filter:** Table OA1.2 shows that, for 29.2% of events, ACLED does not report precise information on their political or military status.[7] The information filter excludes them from

[6]Geo-precision code 1 (the highest precision level): "If the source reporting indicates a particular town, and coordinates are available for that town, the highest precision level, Geo-precision code 1, is recorded". (p35, ACLED, 2023) Geo-precision code 2: "If the source material indicates that activity took place in a small part of a region, and mentions a general area, the event is coded to a town with geo-referenced coordinates to represent that area, and the Geo-precision code 2 is recorded. If activity occurs near a town or a city, this same Geo-precision code 2 is employed." (p35, ACLED, 2023) Geo-precision code 3: "If a larger region is mentioned, the closest natural location noted in reporting (like "border area", "forest", or "sea", among others) – or a provincial capital is used if no other information at all is available – is chosen to represent the region, and Geo-precision code 3 is recorded."

[7]Here are the 10 actors with the highest number of events for which we do not have information on the nature: Labor Group (Morocco); Labor Group (Nigeria); Labor Group (South Africa); Labor Group (Tunisia); Oromo Ethnic Group (Ethiopia); Pastoralists (Nigeria); Resistance Committees (Sudan); Students (South Africa); Women (Democratic Republic of Congo)

the analysis in the following manner. First we exclude actors assigned by ACLED to the category "Unidentified". This represents 33502 distinct events covering 161 actors. Second, we exclude actors for which there is no information recorded on their nature (see footnote 4). This represents 71951 distinct events covering 5388 actors. Last, for actors for which the name is "Militia (Pro-Government)" or "Militia (Students)", we amend their name by mentioning the country in which the event is taking place. We construct 9 different actors for "Militia (Students)" and 15 different actors for "Militia (Pro-Government)"; they are involved in 593 distinct events. Note that the number of events is not affected by this relabeling of names. Overall, after this filter, we end-up with 232878 distinct events covering 5486 actors.

**Step 4- Actor selection:** In line with our theoretical model of raiding, we focus our analysis on violence perpetrated by actors classified as "Rebel groups" and "Political militias". These groups are indeed the most likely to project violence outside their rear base and to perpetrate violence-for-appropriation. Together, rebels and political militias represent 1492 actors (more than one quarter of the total number of actors). In the data, they exert the most lethal forms of violence, namely battles and violence against civilians (see Table OA1.3), participating in 38% of events but responsible for 68% of the total fatalities recorded in the step 3 sample.[8] The *actors selection* filter (step 4) restricts the analysis to these two categories. The resulting sample is made of 1492 actors participating in 87578 events resulting in 379314 fatalities.

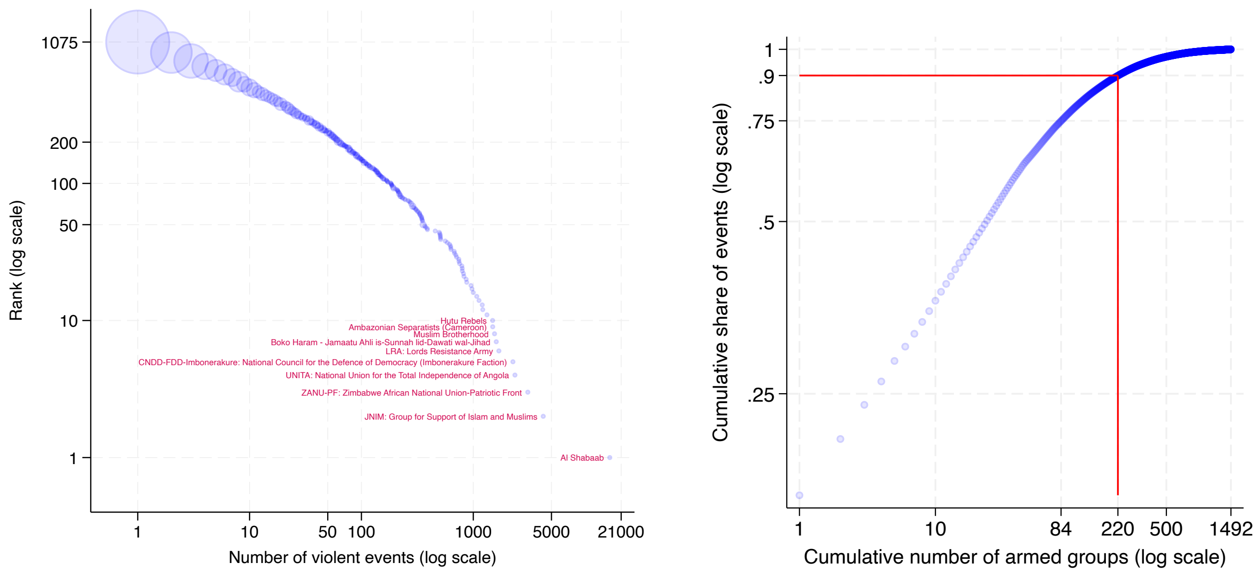
**General Statistics** The violence data feature a granularity that has not been previously documented in the literature. This point is illustrated in Figure OA1.1: Panel (a) reports the log rank-log size relationship in the sample of violent actors at step 4 of our data construction procedure, highlighting the ten major actors, while Panel (b) displays a Lorenz curve for the same data, plotting cumulative counts of actors (ranked by size) against their cumulative share of events. The distribution is highly skewed, as the 220 most violent actors exert over 90% of total violence (80619 events) leading to 362672 fatalities.

[8] At step 3, ACLED records a total of 553745 fatalities associated with 232878 events. Among those, rebel groups and political militias are responsible for 379314 fatalities in 87578 distinct events. Hence, the average number of fatalities per event for rebel groups and militias is 4.3. In contrast, for the rest of the sample, it is 1.2 $\left(\frac{553745-379314}{232878-87578}\right)$. This indicates that events involving rebel groups and political militias tend to be more deadly.

Figure OA1.1: Violence originating from ACLED violent actors

(a) Log rank-log size

(b) Lorenz curve




Note: Both panels are based on the sample of step 4 in Table OA1.1. It includes 1492 actors participating in 87578 violent events. Panel (a) reports the log rank-log size relationship, with the ten major actors being highlighted. The size of the circle represents the number of actors at a given rank. There are 418 actors with only 1 violent event (rank 1075). The actor Al-Shabaab is the most active with 16609 events. Panel (b) reports a Lorenz curve plotting cumulative counts of actors (ranked by size) against cumulative share of events.

## OA2 Anecdotal evidence: recruitment from ethnic groups

We present a range of anecdotal evidence supporting the assumption that an armed group's credible recruitment base originates from its connections to ethnic groups. Interestingly, this holds true not only for small rebel groups or militias but also for large armed group organizations. In 2016, the BBC reported that "[Boko Haram] draws its fighters mainly from the Kanuri ethnic group", while in Uganda the predominantly Muslim Baganda and Basoga ethnic groups make significant contributions to the ranks of the Ugandan ADF fighters[9] and in the Democratic Republic of Congo, the M23 rebel group is "made up primarily of ethnic Tutsis".[10] Often, recruitment aligns with the ethnicity of the leaders. For instance, the leadership of the Revolutionary United Front is primarily drawn from the Temne ethnic group, thus influencing the composition of the troops.[11] The same observation is made for the Resistência Nacional Moçambicana (RENAMO): many leaders are from the Ndau ethnic group (Nuvunga and Adalima, 2001; Jentzsch, 2021). Finally, the leadership of UNITA predominantly consisted of members from Angola's majority Ovimbundu ethnic group, reflecting ethnic favoritism within the organization.[12] Remarkably, even groups operating across multiple countries still recruit along ethnic lines. For instance, reports indicate that the membership of Al-Qaeda primarily hails from Algerian and local Saharan communities, including the Tuaregs and Berabiche tribal clans of Mali, as well as Moroccans residing in urban areas.[13] Similarly, in Mali, tribal affiliations significantly shape the composition of Ansar Dine, with a notable presence from the Ifoghas tribe, highlighting a consistent pattern of ethnic alignment (Maïga, 2016). Furthermore, ethnic recruitment can also result in division and the splintering of armed groups, underscoring the significance of the ethnic dimension. For example, the Liberians United for Reconciliation and Democracy (LURD) initially included members from both the Mandingo and Krahn ethnic groups but eventually divided into two factions due to ethnic tensions. The Movement for Democracy in Liberia (MODEL) emerged predominantly composed of Krahn members, while the Mandingo contingent remained affiliated with LURD.[14] Similarly, Ansar Dine primarily consists of members from the Ifoghas tribe, who initially split from the MIA (Mouvement islamique de l'Azawad) and later from the HCUA (Mouvement islamique de l'Azawad) (Maïga, 2016). Ultimately, many Renamo leaders originated from the Ndau ethnic group; however, leadership later diversified, and ethnic tensions diminished (Jentzsch, 2021). Some anecdotal evidence highlights that recruitment is solely based on ethnic grievances. For instance, Kamwina Nsapu Militia members are mostly from Luba ethnic group and selectively killed non-Luba.[15] One can find many other anectodal evidence: the core of Rally for Congolese Democracy (RCD) is composed of Banaymulenge peo-

[9] https://allafrica.com/stories/200001040079.html

[10] https://www.cfr.org/global-conflict-tracker/conflict/violence-democratic-republic-congo

[11] https://www.refworld.org/reference/countryrep/marp/2003/en/46214

[12] https://en.wikipedia.org/wiki/UNITA

[13] See Cristiani and Fabiani (2011) and https://www.upi.com/Top_News/Special/2011/01/05/Morocco-nabs-members-of-AQIM-cell/UPI-65581294251807/

[14] https://en.wikipedia.org/wiki/Liberians_United_for_Reconciliation_and_Democracy

[15] https://en.wikipedia.org/wiki/Kamwina_Nsapu_rebellion

ple[16]; the National Movement for the Liberation of Azawad is mostly made up of ethnic Tuareg[17]; the Fano Youth Militia ethno-nationalist Amhara militia[18]; and the Front for Patriotic Resistance of Ituri draws soldiers from Ngiti ethnolinguistic group, a subgroup of Lendu[19].

[16] https://en.wikipedia.org/wiki/Rally_for_Congolese_Democracy
[17] https://www.jeuneafrique.com/32589/politique/nord-du-mali-de-l-irr-dentisme-touareg-la-guerre-tribale/
[18] https://en.wikipedia.org/wiki/Fano_(militia)
[19] https://en.wikipedia.org/wiki/Patriotic_Resistance_Front_of_Ituri

## OA3 Matching ACLED actors with Murdock ethnic groups

In this section, we detail the steps 5 and 6 in Table OA1.1. The first step consists of linking ACLED violent actors to Murdock ethnic groups, a task facilitated by the granularity of the violence data (Figure OA1.1). As explained in the main text, this feature enables us to concentrate our analysis on the 220 most violent actors, who are responsible for 90% of the total violence. The second step involves linking the location of each violent event to an ethnic group.

Our methodology delineates origins and destinations based on ethnic homelands covered in the Ethnographic Atlas (Murdock, 1959). This atlas provides insights into the spatial distribution of African ethnicities and compiles various quantitative indicators reflecting the political institutions, cultural practices, and economic characteristics of 1291 societies.[20] This dataset has been extensively used in the literature in economics.[21] Although it has faced some criticism (Wright, 1999; Abad and Maurer, 2021), a recent work by Bahrami-Rad et al. (2021) evaluates the Atlas's validity and concludes that it is a "meaningful source of information".[22]

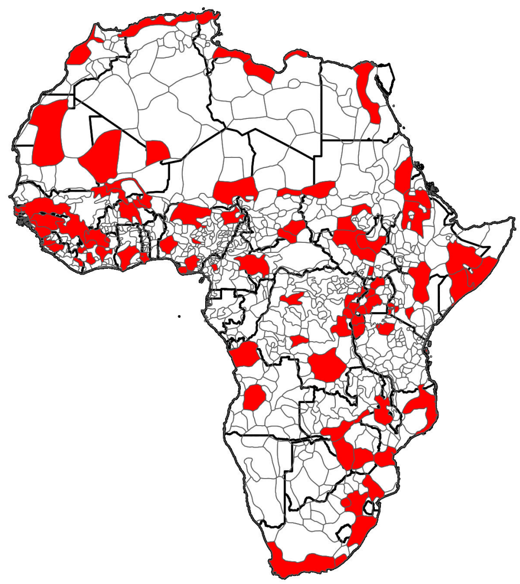
**Step 5- Murdock filter on actors:** To identify the origin of violence, we assign a unique Murdock ethnic group to the 220 violent actors. This task was conducted by the authors and independently cross-validated by a research assistant. The matching procedure involves three main steps: (5a) linking the actors to one or multiple ethnic groups (not necessarily Murdock ethnic groups); (5b) converting these ethnic groups into Murdock ethnic groups; and (5c) selecting a unique Murdock ethnic group among the groups found in the previous steps. At the end, the 220 actors are associated to 87 different Murdock ethnic groups. Figure OA3.2 depicts their spatial distribution.

[20] The dataset was later digitalized by Gray (1999) and the location of those ethnic groups has been downloaded from `https://nathannunn.arts.ubc.ca/data/`

[21] See for instance: Gennaioli and Rainer (2007); Nunn (2008); Nunn and Wantchekon (2011); Michalopoulos and Papaioannou (2013); Berman et al. (2023); Eberle et al. (2024); McGuirk and Nunn (2024).

[22] Specifically, they document positive associations between the historical measures collected by ethnographers and self-reported data from 790,000 individuals across 43 countries.

Figure OA3.2: Murdock ethnic groups matched to actors



Note: Spatial distribution of the 87 Murdock ethnic groups associated to the 220 most violent actors in ACLED (exerting 90% of the observed total violence). The maps displayed in thick black lines the African countries and in grey lines the Murdock ethnic regions.

**5a- Direct and indirect inference:** It is possible to unambiguously link 98 ACLED violent actors to a Murdock ethnic group (*"direct inference"*). For the 122 remaining actors, we combine various informational sources, as outlined below, to identify the ethnic group(s) associated with the actor (*"indirect inference"*):

- *Splits and mergers (33 actors)*. We can identify the ethnic affiliation of several violent actors by simply leveraging the affiliations of other actors. First, 6 actors result from the merger of two other groups (e.g. *Islamic State (West Africa) - Lake Chad Faction and/or Boko Haram - Jamaatu Ahli is-Sunnah lid-Dawati wal-Jihad*). If both actors are linked to the same ethnic group, we assign that group. Otherwise, we assign the ethnic group of the most violent actor. Second, 18 actors are linked to others because they are factions, splinter groups, or were or became part of another group. In this case, we attribute the same ethnic group as the main group to which they are related (see, for instance, *OLF: Oromo Liberation Front (Shane Splinter Faction)*). Third, 3 actors form coalitions (e.g. *CPC: Coalition of Patriots for Change*). In this scenario, we select the ethnic group that forms the majority. If no clear majority exists, we choose the ethnic group of the most violent actors. Finally, 6 actors are related to each other without being classified as a faction but are sufficiently similar to be assigned the same ethnic group (e.g., *FNL: National Forces of Liberation*).

- *Ethnicity of the leader (26 actors)*. In absence of clearcut information, we primarily use information on the ethnicity of its leader to link an ACLED violent actor to an ethnic group. For instance, for *MDC: Movement for Democratic Change (Tsvangirai Faction)*, the leader, Morgan Tsvangirai, is Karanga. Therefore, we link the *MDC* to the Karanga ethnic group.

- *Main ethnic group of the country (37 actors)*. If no information on the leader is available, we check if the ACLED actor is clearly associated with a country; in such cases, we refer to the predominant ethnic group within that country. For instance, for the actor *Islamist Militia (Somalia)*, we use the predominant ethnic group in Somalia, which is Somali.

- *Location of events (16 actors)*. If none of the previous pieces of information are available or relevant, we use external sources that report on the territory of military operations to link a violent actor to an ethnic group. This can be at the regional level, such as the actor *Mayi Mayi Militia (FPP/AP: Popular Patriotic Forces People's Army-Kabido)*, which is primarily active in the Lubero territory, or at the city level, such as *Ansar al-Sharia (Libya)*, which is mainly active in Benghazi and Derna. In the absence of external sources, we project ACLED events onto the Murdock map to determine the distribution of violent events across ethnic groups and retain the most frequently affected one as the ethnic affiliation. For example, most events involving *IM: Islamic Movement* are concentrated in the Murdock region identified as Gwandara. Whenever possible, we prioritize external sources to minimize reliance on the spatial distribution of ACLED events.

- For 9 actors we combine various pieces of information (such as the leader's ethnicity and the ethnic composition of the country) because each piece on its own is uncertain, and multiple sources are needed to verify it.

- For one actor, we directly rely on the information available in the ACD2EPR database because of the lack of other sources.

After this step, 185 actors are associated with only one identified ethnic group, 22 actors with two ethnic groups, 8 actors with three ethnic groups, 3 actors with four ethnic groups and 2 actors with more than four ethnic groups.

**5b- Ethnic conversion:** Following the *direct and indirect inference* procedures, we convert the identified ethnic groups into Murdock ethnic groups. For 43% of the actors, the identified ethnic groups do not match any ethnic group in the Murdock dataset. To address this issue, we use ethnic conversion tables provided by the Linking Ethnic Data from Africa (LEDA) project (Müller-Crepon et al., 2022), which uses a measure of ethnic proximity between groups based on the Ethnologue language tree. This method connects 8100 ethnic categories from eleven databases, including surveys, geographic data, and expert-coded lists. Specifically, we use four of these databases: the Ethnic Power Relations Dataset from the Geographical Research on War, the Unified Platform (Girardin et al., 2015), the Afrobarameter Surveys (`https://www.afrobarometer.org/`), the All Minorities at Risk (`https://cidcm.umd.edu/research/all-minorities-risk-project`) and the Ethnic groups from (Fearon, 2003).[23] However, there are instances where the conversion using LEDA is either not possible or insufficient. In these cases, we rely on credible external sources, which we thoroughly document, to establish a correspondence with Murdock ethnic groups. For actors classified as mergers and factions, we assign the ethnic group(s) of the main group from which they originated (22 actors). The different sources and the number of actors involved are summarized in Table OA3.5:

Table OA3.5: Procedure for converting ethnic group names

| **Sources for the conversion** | **(short) Description** | **Number of actors** |
|---|---|---|
| No conversion | The ethnic group identified is already a Murdock ethnic group. | 125 |
| | Links with other actors | 33 |
| LEDA | The conversion results in a unique Murdock ethnic group. | 21 |
| | The conversion results in multiple Murdock ethnic groups. | 13 |
| External sources | The conversion results in a unique Murdock ethnic group. | 11 |
| | The conversion results in multiple Murdock groups. | 4 |
| LEDA and external sources | LEDA is either not precise enough or provides too many conversions | 7 |
| | Some ethnic groups are in LEDA, while others are not. | 5 |
| No conversion was possible | | 1 |

**5c- Decision rule:** To construct the origins of the bilateral flows of violence between Murdock ethnic regions, we need to assign each violent actor to a *unique* Murdock ethnic group. After the *direct and indirect inference* and *conversion* procedures, we encounter three distinct cases. The first and straightforward case is when these procedures identify a unique Murdock ethnic group associated with the actor (158 actors). Second, some actors are directly assigned the Murdock ethnic group of the primary group with which they are associated, so we retain this Murdock ethnic group (33 actors). Third, the procedure identifies multiple Murdock groups linked to a given actor (29 actors). In such cases, we rely on additional external information to select the most appropriate ethnic group. The various sources that are used are listed below:

- Population (17 actors): Among the ethnic group candidates, we select the ethnic group with

[23] The first three databases are the largest among the eleven available. Additionally, we use the database from Fearon (2003) because it includes a relatively large number of ethnic groups (822 ethnic groups in 160 countries).

the largest population.

- Events (6 actors): We select the ethnic group with the highest number of violent events. For example, *Mayi Mayi Militia* is associated to the Konjo and Hunde ethnic groups. As the fighting activity of the *Mayi Mayi Militia* is concentrated in the Lubero and Beni regions, where the Konjo is the predominant ethnic group, we attribute the Konjo ethnic group to *Mayi Mayi Militia*.

- Coalitions and factions (3 actors): We make use of information on the ethnic affiliation of actors from which they originate.

- Alternatives sources (3 actors): For these actors, we use both actor-specific information and decision rules. For the actor *Ansaroul Islam*, we use external information on ethnicity to select among the candidates. For the actor *Tigray People's Liberation Front*, we use information from the ACD2EPR database. Finally, for the actor *Janjaweed*, we use information on group's leader.

**5d- Validation exercises of the matching procedure:** To assess the quality of our matching procedure, we conduct two validation exercises. The first involves a research assistant who independently processed the data for a random subsample of 157 out of the 220 actors. In 85% of the cases (i.e., for 133 actors), the research assistant identifies at least one Murdock ethnic group that exactly match the one found by the authors. The second validation exercise is based on the ACD2EPR 2021 dataset (Wucherpfennig et al., 2012), which links UCDP Armed Conflict Data (v. 20.1) to EPR-Core 2021 groups.[24,25] This dataset identifies all politically relevant ethnic groups worldwide from 1946 to 2021 and includes data on over 800 groups, coding their access to state power. Among the 220 actors for which we collected ethnic group information, 38 actors are also covered in the ACD2EPR dataset and matched by ACD2EPR to a Murdock group.[26] Encouragingly, our procedure assigns the same Murdock ethnic group to 35 out of those 38 actors.

**Step 6- Murdock filter on events:** This filter restricts the dataset to violent events located in Murdock ethnic regions, resulting in the exclusion of 2284 events. Several factors contribute to the exclusion of certain violent events from the dataset. Firstly, the Sinai Peninsula is entirely absent from the Murdock map, leading to the loss of all violent events from that region. Secondly, lakes are depicted on the Murdock map but are not associated with specific ethnic groups, causing events occurring on lakes to be excluded. Thirdly, the Murdock map has more detailed boundaries compared to the ADMIN 0 map, resulting in events along the coastline potentially falling outside the Murdock map boundaries and being excluded.

[24]The Uppsala Conflict Data Program (UCDP) has recorded ongoing violent conflicts since the 1970s and, together with ACLED, is one of the most popular conflict related databases.

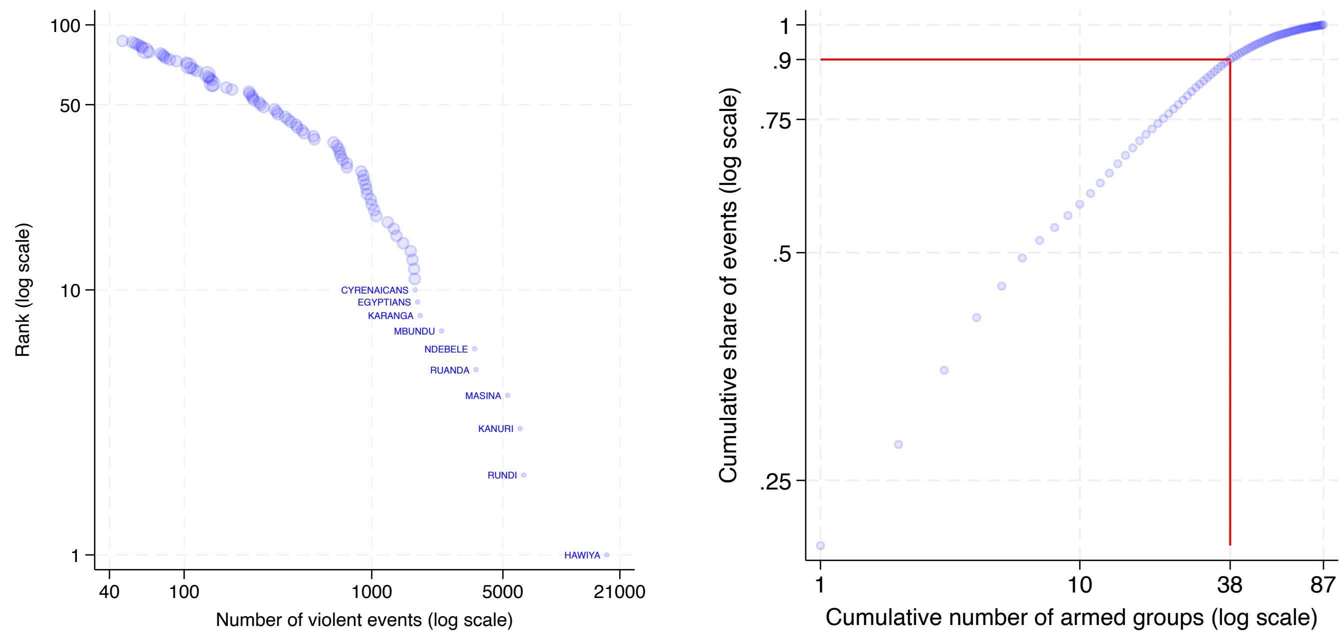
[25]The EPR Core dataset identifies all politically relevant ethnic groups and their access to state power in every country of the world from 1946 to 2021. It includes annual data on over 800 groups and codes the degree to which their representatives held executive-level state power—from total control of the government to overt political discrimination.

[26]Actually 42 armed actors are covered in the ACD2EPR dataset. However, several of them are associated to an ethnic group that is not a Murdock ethnic group. We apply the same methodology used earlier to convert them. For four actors, we could not find a conversion, making comparison impossible. Therefore, out of the 42 actors, we retain only 38 actors for the purpose of the validation exercise.

**General Statistics:** In Figure OA3.3 , we look at the distribution of violence originating from the 87 Murdock ethnic groups (step 6 of our data construction procedure). Panel (a) displays the log rank-log size relationship, highlighting the ten most violent ethnic groups, while panel (b) plots the cumulative counts of ethnic groups (ranked by size) against their cumulative share of events. The distribution is highly skewed, with the 38 most violent actors responsible over 90% of total violence.

Figure OA3.3: Violence originating from Murdock ethnic groups

(a) Log rank-log size (b) Lorenz curve




Note: Both panels are based on the sample of step 6. It includes 78335 events for 87239 observations. Panel (a) reports the log rank-log size relationship, with the ten major ethnic groups being highlighted. Panel (b) reports a Lorenz curve plotting cumulative counts of ethnic groups (ranked by size) against cumulative share of events.

## OA4 Proof of the general equilibrium

The proof of the existence and uniqueness of the equilibrium vector of wages as a fixed-point of equation (14) closely follows Allen (2019).

We start by defining the following function $Z_i(\mathbf{w})$ for $i \in N$ as:

$$Z_i(\mathbf{w}) \equiv \frac{1}{w_i}\left(\sum_n \beta_{in}(\mathbf{w}) w_n \bar{L}_n - w_i \bar{L}_i\right), \tag{OA4.1}$$

where

$$\beta_{in}(\mathbf{w}) \equiv (1-s_n) \times \frac{w_i^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_i}{\xi_{in}}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_k \left(\frac{\psi_k}{\xi_{kn} w_k}\right)^{\frac{\gamma}{1-\gamma}}} + s_n \times \frac{w_i^{1-\sigma}\left(\frac{A_i}{\tau_{in}}\right)^{\sigma-1}}{\sum_k \left(\frac{A_k}{\tau_{kn} w_k}\right)^{\sigma-1}}.$$

If $Z_i(\mathbf{w}) = 0$ for all $i \in N$, then $\mathbf{w}$ represents a price equilibrium. Similarly, when $Z_i(\mathbf{w}) > 0$, it indicates that country $i$ is selling more goods than it is earning. In this context, we can interpret $Z_i(.)$ as the excess demand function for goods originating from country $i \in N$.

For the remaining of the proof, it will turn out to be useful to write $Z_i(.)$ as:

$$Z_i(\mathbf{w}) = Z_i^F(\mathbf{w}) + Z_i^P(\mathbf{w}), \tag{OA4.2}$$

where

$$Z_i^F(\mathbf{w}) = \frac{1}{w_i}\left[\sum_n \frac{w_i^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_i}{\xi_{in}}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_k \left(\frac{\psi_k}{\xi_{kn} w_k}\right)^{\frac{\gamma}{1-\gamma}}}(1-s_n) w_n \bar{L}_n - (1-s_i) w_i \bar{L}_i\right]$$

$$Z_i^P(\mathbf{w}) = \frac{1}{w_i}\left[\sum_n \frac{\left(\frac{A_i}{\tau_{in} w_i}\right)^{\sigma-1}}{\sum_k \left(\frac{A_k}{\tau_{kn} w_k}\right)^{\sigma-1}} s_n w_n \bar{L}_n - s_i w_i \bar{L}_i\right].$$

Indeed:

$$\begin{aligned}
w_i Z_i(\mathbf{w}) &= \sum_n \beta_{in}(\mathbf{w}) w_n \bar{L}_n - w_i \bar{L}_i \\
&= \sum_n \beta_{in}(\mathbf{w}) w_n \bar{L}_n - w_i \bar{L}_i + s_i w_i \bar{L}_i - s_i w_i \bar{L}_i \\
&= \sum_n \beta_{in}(\mathbf{w}) w_n \bar{L}_n - (1-s_i) w_i \bar{L}_i - s_i w_i \bar{L}_i \\
&= \sum_n (1-s_n) \times \frac{w_i^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_i}{\xi_{in}}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_k \left(\frac{\psi_k}{\xi_{kn} w_k}\right)^{\frac{\gamma}{1-\gamma}}} - (1-s_i) w_i \bar{L}_i + \sum_n s_n \times \frac{w_i^{1-\sigma}\left(\frac{A_i}{\tau_{in}}\right)^{\sigma-1}}{\sum_k \left(\frac{A_k}{\tau_{kn} w_k}\right)^{\sigma-1}} - s_i w_i \bar{L}_i.
\end{aligned}$$

Hence:

$$Z_i(\mathbf{w}) = \frac{1}{w_i}\left[\sum_n (1-s_n) \times \frac{w_i^{-\frac{\gamma}{1-\gamma}}(\frac{\psi_i}{\xi_{in}})^{\frac{\gamma}{1-\gamma}}}{\sum_k (\frac{\psi_k}{\xi_{kn} w_k})^{\frac{\gamma}{1-\gamma}}} - (1-s_i) w_i \bar{L}_i\right] + \frac{1}{w_i}\left[\sum_n s_n \times \frac{w_i^{1-\sigma}\left(\frac{A_i}{\tau_{in}}\right)^{\sigma-1}}{\sum_k \left(\frac{A_k}{\tau_{kn} w_k}\right)^{\sigma-1}} - s_i w_i \bar{L}_i\right]$$
$$= Z_i^F(\mathbf{w}) + Z_i^P(\mathbf{w}).$$

**Existence**

$Z_i(\mathbf{w})$ has the following properties:

$(i)$ For all $\mathbf{w} \gg 0$ (ie, for all $\mathbf{w}$ such that $w_i > 0$ for all $i \in N$) and for all $i \in N$, $Z_i(.)$ is continuous. This is immediate from the definition of $Z_i(\mathbf{w})$.

$(ii)$ For all $i \in N$, $Z_i(.)$ is homogeneous of degree zero. Indeed, for any $\lambda > 0$:

$$\begin{aligned} Z_i(\lambda \mathbf{w}) =& \frac{1}{\lambda w_i}\left(\sum_n \frac{(\frac{\psi_i}{\xi_{in}\lambda w_i})^{\frac{\gamma}{1-\gamma}}}{\sum_k (\frac{\psi_k}{\xi_{kn}\lambda w_k})^{\frac{\gamma}{1-\gamma}}}(1-s_n)\lambda w_n \bar{L}_n - (1-s_i)\lambda w_i \bar{L}_i + \sum_n \frac{(\frac{A_i}{\tau_{in}\lambda w_i})^{\sigma-1}}{\sum_k (\frac{\psi_k}{\xi_{kn}\lambda w_k})^{\sigma-1}} s_n \lambda w_n \bar{L}_n \right.\\ &\left. - s_i \lambda w_i \bar{L}_i\right)\\ =& \frac{1}{w_i}\left(\sum_n \frac{\lambda^{\frac{-\gamma}{1-\gamma}}(\frac{\psi_i}{\xi_{in} w_i})^{\frac{\gamma}{1-\gamma}}}{\lambda^{\frac{-\gamma}{1-\gamma}}\sum_k (\frac{\psi_k}{\xi_{kn} w_k})^{\frac{\gamma}{1-\gamma}}}\frac{1}{\lambda}(1-s_n)\lambda w_n \bar{L}_n - \frac{1}{\lambda}(1-s_i)\lambda w_i \bar{L}_i\right.\\ &\left. + \sum_n \frac{\lambda^{1-\sigma}(\frac{A_i}{\tau_{in} w_i})^{\sigma-1}}{\sum_k \lambda^{1-\sigma}(\frac{A_k}{\tau_{kn} w_k})^{\sigma-1}}\frac{1}{\lambda} s_n \lambda w_n \bar{L}_n - \frac{1}{\lambda} s_i \lambda w_i \bar{L}_i\right)\\ =& Z_i(\mathbf{w}). \end{aligned}$$

$(iii)$ For all $\mathbf{w} \gg 0$, we have:

$$\sum_{i \in N} w_i Z_i(\mathbf{w}) = 0.$$

Indeed:

$$\sum_i w_i Z_i(\mathbf{w}) = \sum_i \left(\sum_n \beta_{in}(\mathbf{w}) w_n \bar{L}_n - w_i \bar{L}_i\right).$$

Noticing that:

$$w_i l_i + w_i L_i = \sum_n \frac{\xi_{in}^{-\frac{\gamma}{1-\gamma}}\left(\frac{\psi_i}{w_i}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_k \xi_{kn}^{-\frac{\gamma}{1-\gamma}}\left(\frac{\psi_k}{w_k}\right)^{\frac{\gamma}{1-\gamma}}}(1-s_n) w_n \bar{L}_n + \sum_n \frac{\tau_{in}^{-(\sigma-1)}\left(\frac{A_i}{w_i}\right)^{\sigma-1}}{\sum_k \tau_{kn}^{-(\sigma-1)}\left(\frac{A_k}{w_k}\right)^{\sigma-1}} s_n w_n \bar{L}_n$$
$$\Leftrightarrow w_i \bar{L}_i = \sum_n \beta_{in}(\mathbf{w}) w_n \bar{L}_n.$$

It follows that:

$$\sum_i w_i Z_i(\mathbf{w}) = \sum_i (w_i \bar{L}_i - w_i \bar{L}_i) = 0.$$

$(iv)$ For all $\mathbf{w} \gg 0$, it exists a $k > 0$ such that $Z_i(\mathbf{w}) > -k$ for all $i \in N$. Indeed, the lower bound on $Z_i(\mathbf{w})$ is implied by:

$$Z_i(\mathbf{w}) = \frac{1}{w_i}\left(\sum_n \beta_{in}(\mathbf{w}) w_n \bar{L}_n - w_i \bar{L}_i\right) = \underbrace{\frac{1}{w_i}\left(\sum_n \beta_{in}(\mathbf{w}) w_n \bar{L}_n\right)}_{>0} - \bar{L}_i > -\bar{L}_i.$$

Hence, if we define $k \equiv \max_{i \in S} \bar{L}_i$, it follows that $Z_i(\mathbf{w}) > -k \; \forall i \in N$.

$(v)$ Consider any $\mathbf{w} \in \mathbb{R}^{\|N\|}$ such that there exists a $l \in N$ where $w_l = 0$ and an $l' \in N$ where $w_{l'} > 0$. Consider any sequence of wages such that $\mathbf{w}^m \to \mathbf{w}$ as $m \to \infty$. Then:

$$\max_{i \in N} Z_i(\mathbf{w}^m) \to \infty.$$

To show it, we demonstrate that this result holds for $Z_i^F(\mathbf{w})$ and $Z_i^P(\mathbf{w})$. By addivity, the result will hold for $Z_i(\mathbf{w})$.

For readibility , we define: $K(i,n,k) = \frac{w_i^{1-\sigma} \times (\frac{A_i}{\tau_{in}})^{\sigma-1}}{\sum_k (\frac{A_k}{\tau_{kn} w_k})^{\sigma-1}}$

To begin with, note that:

$$\max_{i \in N} Z_i^P(\mathbf{w}^m) = \max_{i \in N}\left\{\frac{1}{w_i}\sum_n K(i,n,k) s_n w_n \bar{L}_n - s_i \bar{L}_i\right\}.$$

Then, we have:

$$\max_{i \in N}\left\{\frac{1}{w_i}\sum_n K(i,n,k) s_n w_n \bar{L}_n - s_i \bar{L}_i\right\} > \max_{i \in N}\left\{\frac{1}{w_i}\right\} \max_{i \in N}\left\{\sum_n K(i,n,k) s_n w_n \bar{L}_n - s_i \bar{L}_i\right\}.$$

Then note:

$$\max_{i \in N}\left\{\frac{1}{w_i}\right\} \max_{i \in N}\left\{\sum_n K(i,n,k) s_n w_n \bar{L}_n - s_i \bar{L}_i\right\} > \max_{i \in N}\left\{\frac{1}{w_i}\right\} \max_{i \in N}\left\{\sum_n K(i,n,k) s_n w_n \bar{L}_n\right\} - \max_{i \in N}\left\{s_i \bar{L}_i\right\}.$$

Finally:

$$\max_{i\in N}\left\{\frac{1}{w_i}\right\}\max_{i\in N}\left\{\sum_n K(i,n,k)s_n w_n \bar{L}_n\right\}-\max_{i\in N}\left\{s_i\bar{L}_i\right\}>\max_{i\in N}\left\{\frac{1}{w_i}\right\}\max_{n\in N}\left\{K(i,n,k)s_n w_n\bar{L}_n\right\}-\max_{i\in N}\left\{s_i\bar{L}_i\right\}$$
$$\Rightarrow \max_{i\in N}\left\{\frac{1}{w_i}\right\}\max_{i\in N}\left\{\sum_n K(i,n,k)s_n w_n \bar{L}_n\right\}-\max_{i\in N}\left\{s_i\bar{L}_i\right\}>\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)s_n\bar{L}_n\right\}-\max_{i\in N}\left\{s_i\bar{L}_i\right\}$$
$$\Leftrightarrow \max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)s_n\bar{L}_n\right\}-\max_{i\in N}\left\{s_i\bar{L}_i\right\}>\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)s_n\bar{L}_n\right\}-\max_{i\in N}\left\{\bar{L}_i\right\},$$

since $s_i\in[0,1]$. Hence:

$$\max_{i\in N} Z_i^P(\mathbf{w}^m)=\max_{i\in N}\left\{\frac{1}{w_i}\sum_n K(i,n,k)s_n w_n\bar{L}_n-s_i\bar{L}_i\right\}>\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)s_n\bar{L}_n\right\}-\max_{i\in N}\left\{\bar{L}_i\right\}.$$

Since $\max_{i\in N}\bar{L}_i$ is finite, it is sufficient to show that $\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)s_n\bar{L}_n\right\}\to\infty$ to demonstrate that $\max_{i\in N} Z_i^P(\mathbf{w}^m)\to\infty$ according to the above inequality.
Firstly, note that:

$$\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)s_n\bar{L}_n\right\}>\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)s_n\right\}\max_{l\in N}\left\{\bar{L}_n\right\}.$$

We now use the maximization across $n$ to choose the highest wage taking into account $s_n$:

$$\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)s_n\bar{L}_n\right\}>\max_{l\in N}\left\{\bar{L}_n\right\}\max_{n\in N}\left\{w_n s_n\right\}\max_{i\in N}\left\{\frac{1}{w_i}K(i,n,k)\right\}.$$

Note that:

$$\frac{1}{w_i}K(i,n,k)=\frac{w_i^{-\sigma}\times(\frac{A_i}{\tau_{in}})^{\sigma-1}}{\sum_k(\frac{A_k}{\tau_{kn}w_k})^{\sigma-1}}.$$

We now make $\sum_k(\frac{A_k}{\tau_{kn}w_k})^{\sigma-1}=\sum_k(\frac{A_k}{\tau_{kn}})^{\sigma-1}w_k^{1-\sigma}$ as large as possible. Recall that $\sigma>1$ so $1-\sigma<0$. Hence, to make this expression as large as possible, it is sufficient to make $w_k^{1-\sigma}=\frac{1}{w_k^{\sigma-1}}$ as large as possible. To do so, we need to make $w_k$ as small as possible. Hence, set $w_k=\min_{l\in N}\left\{w_l\right\}$. We then have:

$$\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)s_n\bar{L}_n\right\}>C\times\max_{i\in N}\frac{w_i^{-\sigma}\times(\frac{A_i}{\tau_{in}})^{\sigma-1}}{\sum_k(\frac{A_k}{\tau_{kn}})^{\sigma-1}\Big(\min_{l\in N}\{w_l\}\Big)^{1-\sigma}},$$

where $C=\max_{l\in N}\{\bar{L}_l\}\max_{n\in N}\{w_n s_n\}$.
We now use the maximization across $i$ to make the numerator as large as possible (taking into

account $-\sigma < -1$):

$$\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)s_n\bar{L}_n\right\} > C \times \max_{i\in N}\frac{\Big(\min_{l\in N}\{w_l\}\Big)^{-\sigma}\times(\frac{A_i}{\tau_{in}})^{\sigma-1}}{\sum_k(\frac{A_k}{\tau_{kn}})^{\sigma-1}\Big(\min_{l\in N}\{w_l\}\Big)^{1-\sigma}}.$$

Cancelling like terms:

$$\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)s_n\bar{L}_n\right\} > C \times \max_{i\in N}\frac{(\frac{A_i}{\tau_{in}})^{\sigma-1}}{\sum_k(\frac{A_k}{\tau_{kn}})^{\sigma-1}}\times\Big(\min_{l\in N}\{w_l\}\Big)^{-1}.$$

Since there exists an $l \in N$ such that $w_l^m \to 0$ as $m \to \infty$, it follows that:

$$C \times \max_{i\in N}\frac{(\frac{A_i}{\tau_{in}})^{\sigma-1}}{\sum_k(\frac{A_k}{\tau_{kn}})^{\sigma-1}}\times\Big(\min_{l\in N}\{w_l\}\Big)^{-1} \to \infty,$$

implying that:

$$\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)s_n\bar{L}_n\right\} \to \infty,$$

and hence:

$$\max_{i\in N} Z_i^P(\mathbf{w}) \to \infty.$$

The reasonning is exactly the same for $Z_i^F(\mathbf{w})$. Define:

$$K(i,n,k) = \frac{w_i^{-\frac{\gamma}{1-\gamma}}(\frac{\psi_i}{\xi_{in}})^{\frac{\gamma}{1-\gamma}}}{\sum_k(\frac{\psi_k}{\xi_{kn}w_k})^{\frac{\gamma}{1-\gamma}}}.$$

Hence:

$$\max_{i\in N} Z_i^F(\mathbf{w}^m) = \max_{i\in N}\left\{\frac{1}{w_i}\sum_n K(i,n,k)(1-s_n)w_n\bar{L}_n - (1-s_i)\bar{L}_i\right\}.$$

Then, we have:

$$\begin{aligned}
&\max_{i\in N} Z_i^F(\mathbf{w}^m) > \max_{i\in N}\left\{\frac{1}{w_i}\right\}\max_{n\in N}\left\{K(i,n,k)(1-s_n)w_n\bar{L}_n\right\} - \max_{i\in N}\left\{(1-s_i)\bar{L}_i\right\} \\
&\Longrightarrow \max_{i\in N} Z_i^F(\mathbf{w}^m) > \max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)(1-s_n)\bar{L}_n\right\} - \max_{i\in N}\left\{(1-s_i)\bar{L}_i\right\} \\
&\Longrightarrow \max_{i\in N} Z_i^F(\mathbf{w}^m) > \max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)(1-s_n)\bar{L}_n\right\} - \max_{i\in N}\left\{\bar{L}_i\right\}.
\end{aligned}$$

Since $\max_{i\in N}\bar{L}_i$ is finite, it is sufficient to show that $\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)(1-s_n)\bar{L}_n\right\} \to \infty$ to

demonstrate that $\max_{i\in N} Z_i^F(\mathbf{w}^m) \to \infty$ according to the above inequality.
Firstly, note that:

$$\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)(1-s_n)\bar{L}_n\right\} > \max_{l\in N}\left\{\bar{L}_n\right\}\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)(1-s_n)\right\}.$$

We now use the maximization across $n$ to choose the highest wage taking into account $s_n$:

$$\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)(1-s_n)\bar{L}_n\right\} > \max_{l\in N}\left\{\bar{L}_n\right\}\max_{n\in N}\left\{w_n(1-s_n)\right\}\max_{i\in N}\left\{\frac{1}{w_i}K(i,n,k)\right\}.$$

Note that:

$$\frac{1}{w_i}K(i,n,k) = \frac{w_i^{-\frac{1}{1-\gamma}}\left(\frac{\psi_i}{\xi_{in}}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_k\left(\frac{\psi_k}{\xi_{kn}}\right)^{\frac{\gamma}{1-\gamma}}w_k^{\frac{-\gamma}{1-\gamma}}}.$$

We now make $\sum_k\left(\frac{\psi_k}{\xi_{kn}}\right)^{\frac{\gamma}{1-\gamma}}w_k^{\frac{-\gamma}{1-\gamma}}$ as large as possible. Since $\gamma < 1$, we have $\frac{-\gamma}{1-\gamma} < 0$ and so to make the above expression as large as possible, it is sufficient to set $w_k = \min_{l\in N}\left\{w_l\right\}$. We then have:

$$\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)(1-s_n)\bar{L}_n\right\} > C\times\max_{i\in N}\frac{w_i^{-\frac{1}{1-\gamma}}\left(\frac{\psi_i}{\xi_{in}}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_k\left(\frac{\psi_k}{\xi_{kn}}\right)^{\frac{\gamma}{1-\gamma}}\left(\min_{l\in N}\{w_l\}\right)^{\frac{-\gamma}{1-\gamma}}},$$

where $C = \max_{l\in N}\{\bar{L}_l\}\max_{n\in N}\{w_n(1-s_n)\}$.
We now use the maximization across $i$ to make the numerator as large as possible:

$$\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)(1-s_n)\bar{L}_n\right\} > C\times\max_{i\in N}\frac{\left(\min_{l\in N}\{w_l\}\right)^{-\frac{1}{1-\gamma}}\left(\frac{\psi_i}{\xi_{in}}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_k\left(\frac{\psi_k}{\xi_{kn}}\right)^{\frac{\gamma}{1-\gamma}}\left(\min_{l\in N}\{w_l\}\right)^{\frac{-\gamma}{1-\gamma}}}.$$

Cancelling like terms:

$$\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)(1-s_n)\bar{L}_n\right\} > C\times\max_{i\in N}\frac{\left(\frac{\psi_i}{\xi_{in}}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_k\left(\frac{\psi_k}{\xi_{kn}}\right)^{\frac{\gamma}{1-\gamma}}}\times\left(\min_{l\in N}\{w_l\}\right)^{-1}.$$

Since there exists an $l \in N$ such that $w_l^m \to 0$ as $m \to \infty$, it follows that:

$$C\times\max_{i\in N}\frac{\left(\frac{\psi_i}{\xi_{in}}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_k\left(\frac{\psi_k}{\xi_{kn}}\right)^{\frac{\gamma}{1-\gamma}}}\times\left(\min_{l\in N}\{w_l\}\right)^{-1} \to \infty,$$

implying that:

$$\max_{i,n\in N}\left\{\frac{w_n}{w_i}K(i,n,k)(1-s_n)\bar{L}_n\right\} \to \infty,$$

and hence:

$$\max_{i \in N} Z_i^F(\mathbf{w}) \to \infty.$$

Since $Z_i(\mathbf{w}) = Z_i^P(\mathbf{w}) + Z_i^F(\mathbf{w})$, it follows that: $\max_{i \in N} Z_i(\mathbf{w}) \to \infty$.

We achieve the proof of existence by making use of the following theorem:

**Theorem 1.** *If a function $\{Z_i(.)\}_{i \in N}$ satisfies conditions $(i)$ to $(v)$, then there exists a $\boldsymbol{w}^* \gg 0$ such that $Z_i(\boldsymbol{w}^*) = 0 \ \forall i \in N$. (This theorem comes directly from Mas-Colell et al. (1995) Proposition 17.C.1 on p.585.)*

**Uniqueness**

A differentiable excess demand function $\{Z_i(.)\}_{i \in N}$ is said to satisfy the gross substitute property if $\forall i \in N$:

$$\frac{\partial Z_i(\mathbf{w})}{\partial w_j} > 0 \quad \forall j \neq i,$$

We show that $\{Z_i(\mathbf{w})\}_{i \in N}$ follows the gross substitute property. We start by differentiating $Z_i^P(\mathbf{w})$ w.r.t. $w_j$:

$$\begin{aligned}
\frac{\partial Z_i^F(\mathbf{w})}{\partial w_j} =& \frac{\partial}{\partial w_j} \frac{1}{w_i} \left( \sum_n \frac{(\frac{\psi_i}{\xi_{in} w_i})^{\frac{\gamma}{1-\gamma}}}{\sum_k (\frac{\psi_k}{\xi_{kn} w_k})^{\frac{\gamma}{1-\gamma}}} (1 - s_n) w_n \bar{L}_n - (1 - s_i) w_i \bar{L}_i \right) \iff \\
=& \frac{1}{w_i} \left[ \frac{(\frac{\psi_i}{\xi_{ij} w_i})^{\frac{\gamma}{1-\gamma}} (1 - s_j) \bar{L}_j}{\sum_k (\frac{\psi_k}{\xi_{kj} w_k})^{\frac{\gamma}{1-\gamma}}} + \sum_k \frac{(\frac{\psi_i}{\xi_{ik} w_i})^{\frac{\gamma}{1-\gamma}} (1 - s_k) w_k \bar{L}_k (\frac{\psi_j}{\xi_{jk}})^{\frac{\gamma}{1-\gamma}} (-\frac{\gamma}{1-\gamma})(-1) w_j^{\frac{-\gamma}{1-\gamma} - 1}}{\left( \sum_j (\frac{\psi_j}{\xi_{jk} w_j})^{\frac{\gamma}{1-\gamma}} \right)^2} \right] > 0.
\end{aligned}$$

since the first term in the parentheses is positive and the second term is also positive since $\gamma < 1$. We do the same for $Z_i^P(\mathbf{w})$:

$$\frac{\partial Z_i^P(\mathbf{w})}{\partial w_j} = \frac{1}{w_i} \left[ \frac{(\frac{A_i}{\tau_{ij} w_i})^{\sigma-1} s_j \bar{L}_j}{\sum_k (\frac{A_k}{\tau_{kj} w_k})^{\sigma-1}} + \sum_k \frac{(\frac{A_i}{\tau_{ik} w_i})^{\sigma-1} s_k \bar{L}_k w_k (\frac{A_i}{\tau_{jk}})^{\sigma-1} (-1)(1-\sigma) w_j^{-\sigma}}{\left( \sum_j (\frac{A_j}{\tau_{jk} w_j})^{\sigma-1} \right)^2} \right] > 0.$$

since the first term in the parentheses is positive and the second term is also positive because $\sigma > 1$. Thus it follows that, $\forall i \in N$:

$$\frac{\partial Z_i^F(\mathbf{w})}{\partial w_j} + \frac{\partial Z_i^P(\mathbf{w})}{\partial w_j} > 0 \Leftrightarrow \frac{\partial Z_i(\mathbf{w})}{\partial w_j} > 0 \quad \forall j \neq i$$

We now use the following theorem to conclude the proof:

**Theorem 2.** *If a function $\{Z_i(.)\}_{i \in N}$ satisfies the gross substitute property and is homogeneous of degree zero, then the equilibrium $\boldsymbol{w}^* \gg 0$ such that $Z_i(\boldsymbol{w}^*) = 0 \ \forall i \in N$ is unique (to scale). (This theorem comes directly from Mas-Colell et al. (1995) Proposition 17.F.3 on p.613. )*

# References


Abad, L. A. and N. Maurer (2021). History never really says goodbye: A critical review of the persistence literature. *Journal of Historical Political Economy 1*(1), 31–68.

Allen, T. (2019). Modern spatial economics. *Lecture given at the University of Zurich*.

Bahrami-Rad, D., A. Becker, and J. Henrich (2021). Tabulated nonsense? testing the validity of the ethnographic atlas. *Economics Letters 204*, 109880.

Berman, N., M. Couttenier, and V. Girard (2023). Natural resources and ethnic identity. *The Economic Journal 133*(653), 705–1737.

Cristiani, D. and R. Fabiani (2011). Al qaeda in the islamic maghreb (aqim): Implications for algeria's regional and international relations. *Istituto Affari Internazionali (IAI) Working Papers 11, no. 7*.

Eberle, U., D. Rohner, and M. Thoenig (2024). Heat and hate: Climate security and farmer-herder conflicts in africa. (15542).

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth 8*(2), 195–222.

Gennaioli, N. and I. Rainer (2007). The modern impact of precolonial centralization in africa. *Journal of Economic Growth 12*(3), 185–234.

Girardin, L., P. Hunziker, L.-E. Cederman, N.-C. Bormann, S. Rüegger, and M. Vogt (2015). Growup - geographical research on war, unified platform. *ETH Zurich*.

Gray, P. (1999). A corrected ethnographic atlas. *World Cultures 10(1)*, 24–85.

Jentzsch, C. (2021). Renamo and Mozambique. *Oxford Research Encyclopedia of African History*.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995). *Microeconomic Theory*. New York: Oxford University Press.

Maïga, I. (2016). Armed groups in Mali: Beyond the labels. Technical report.

McGuirk, E. and N. Nunn (2024). Transhumant pastoralism, climate change and conflict in africa. *The Review of Economic Studies Forthcoming*.

Michalopoulos, S. and E. Papaioannou (2013). Pre-colonial ethnic institutions and contemporary african development. *Econometrica 81*(1), 113–152.

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

Murdock, G. P. (1959). An Atlas of African History. JD Fage. *American Anthropologist 61*(3), 530–531.

Nunn, N. (2008). The long-term effects of africa's slave trades. *The Quarterly Journal of Economics 123*(1), 139–176.

Nunn, N. and L. Wantchekon (2011). The slave trade and the origins of mistrust in africa. *American Economic Review 101*(7), 3221–52.

Nuvunga, A. and J. Adalima (2001). *Mozambique Democratic Movement (MDM): an analysis of a new opposition party in Mozambique*. Friedrich Ebert Stiftung.

Wright, D. R. (1999). "what do you mean there were no tribes in africa?": Thoughts on boundaries: And related matters: In precolonial africa. *History in Africa 26*, 409–426.

Wucherpfennig, J., M. Nils W., C. Lars-Erik, and G. Kristian Skrede (2012). Ethnicity, the state, and the duration of civil war. *World Politics 64*(1), 79–115.