SciencesPo
DEPARTMENT OF ECONOMICS

Discussion Paper

THE GRAVITY OF VIOLENCE

Mathieu Couttenier, Julian Marcoux, Thierry Mayer, and Mathias Thoenig

SCIENCES PO ECONOMICS DISCUSSION PAPER

No. 2024-10

THE GRAVITY OF VIOLENCE *

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

September 17, 2024

Abstract

This paper presents a framework for estimating and simulating a quantitative spatial model of trade and violence. In this new theoretical and empirical setup, designed for the analysis of bilateral flows, we first model the general equilibrium interactions between the economic and fighting margins in a micro-founded setup. We then show how the structural parameters can be recovered from the data in a simple and transparent way. A central element of the procedure consists in estimating a structural gravity equation of violence. Studying continental Africa over the period 1997 to 2023, we test the key predictions of the model and uncover new facts related to spatial frictions and conflicts. Finally, the model is simulated to quantify different policy-relevant experiments in weakly institutionalized contexts where insecurity is pervasive. Positive and spatially-diffusing feedback loops between development and pacification are one of the main findings of our analysis.

Keywords: Violence, conflicts, trade, development, gravity.

JEL Codes: F1, D74, O1, R1

*Corresponding author: Mathias Thoenig (mathias.thoenig@unil.ch). Helpful comments from David Atkin, Patrick Francois, Jeremy Laurent-Lucchetti, Fernando Parro, Luigi Pascali, Frederic Robert-Nicoud, Esteban Rossi-Hansberg, as well as from participants of seminars and conferences at Alicante, Bristol, CEMFI, ECONtribute in Cologne, Graduate Institute, Kiel-CEPR Conference on Geoeconomics, Mannheim, Nottingham, NYUAD, Oxford, Polytechnique, Pompeu Fabra, Princeton-Yale-CEPR Symposium on Political Economy, Princeton IES Summer Trade Workshop, Saint-Gallen, SMYE 2023, UAB, Warwick are gratefully acknowledged. We thank Matteo Moglia for excellent research assistance. Mathias Thoenig acknowledges financial support from the SNF grant "Quantitative Spatial Models of Violence and Migration" (204899).

†ENS de Lyon, Center for Economic Research on Governance, Inequality and Conflict, 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 Introduction

Armed conflicts and organized violence remain pervasive in the modern world: approximately one third of the world population lived in a country experiencing a conflict in 2016 (Korovkin and Makarin, 2021). This high prevalence of violence carries significant implications for economic growth, particularly for the development of low-income countries. The Organisation for Economic Co-operation and Development (OECD) reported in 2009 that 60% of the world's poorest countries were affected by armed conflicts, which it considered a major hindrance to their ability to progress and develop (OECD, 2009).

Furthermore, development policies that are thought to work reasonably well in peaceful environments may not necessarily be effective in violent contexts. A recent empirical literature, following Nunn and Qian (2014), has shown that humanitarian aid and food assistance programs may inadvertently exacerbate violence in recipient countries. Additionally, Crost et al. (2014) document instances in the Philippines where insurgents deliberately undermined a development program funded by the World Bank. Their motivation stemmed from the belief that the potential success of the program could weaken their support among the local population. These findings underscore the need for a more comprehensive quantitative assessment of the functioning of development policies in conflict-prone countries.[1] One illustrative example of the complexity involved can be found in the ambiguous impact of road and infrastructure construction: on the one hand, it can enhance trade and appease tensions, while on the other hand, it may also facilitate border-crossing and intensify violence intended to grab that newly generated income. Similarly, structural transformation can amplify rural-urban disparities, thereby rendering specific regions more vulnerable to conflicts. In order to design effective policy interventions in the most fragile regions, it is therefore essential to comprehend the intricate general equilibrium interactions between economic factors and conflicts.[2]

The objective of this paper is to develop a quantitative spatial model that combines the economic and fighting margins in a tractable and estimable manner. Our approach, focused on the African continent, is based on two premises. First, violence in Africa involves groups with a remarkable ability for spatial projection. These groups often extend the reach of their violence far beyond their rear bases, enabling us to measure the violence as a flow with discernible origins and destinations: identical when violence is exerted locally, origin and destination do differ when violence is exerted at a distance. This perspective departs from the "place-based" analysis prevailing in existing empirical work on intrastate conflict data and allows for a more subtle understanding of the nature of violence. Second, in the tradition of the influential works of Olson (1993), Tilly (1985) and Dal Bó and Dal Bó (2011), we model conflict and war-making as organized crime: fighting

[1] Recognizing this imperative, the United Nations has responded by adopting a Triple Nexus Development-Peace-Humanitarian doctrine aimed at addressing this concern: https://www.un.org/humansecurity/wp-content/uploads/2022/03/FINAL-Triple-Nexus-Guidance-Note-for-web_compressed.pdf

[2] Quoting McGuirk and Burke (2020) "Modeling general equilibrium forces is important as economic shocks that alter the opportunity cost of violence could also affect the spoils of victory or a government's capacity to repel insurgents, yielding an unclear relationship. This ambiguity is reflected in a markedly inconclusive empirical literature, characterized by inconsistent findings and by significant identification challenges: income may affect conflict; conflict may affect income; and both".

groups are primarily seen as stationary or roving bandits exerting violence to appropriate population's income. This view has received considerable empirical support in the literature.[3]

A modeling challenge relates to the coupling of the trade and fighting decision margins in general equilibrium. To tackle this issue, we use the fact that both trade and conflict can be microfounded in models that generate CES-like functional forms governing aggregate behaviors (referred to as gravity in trade models, and as contest success function in conflict models). This implies that we can use recent advances in methods proposed by the quantitative spatial economics literature (Redding and Rossi-Hansberg, 2017; Redding, 2022) to model the conflict margin. Of particular importance is our theoretical derivation of a structural "gravity of violence", parallel to the trade-in-goods counterpart, a major building block for quantitative spatial models (Allen et al., 2020). The gravity of violence equation embeds the standard channels through which income affects violence, which have been the focus of the previous literature. It also reveals a new channel related to *spatial competition* in the appropriation market among fighting groups, which we refer to as the Multilateral Resistance of Violence (MRV), using the terminology introduced in trade by Anderson and van Wincoop (2003). Then, we show how the structural parameters can be recovered from the data in a simple and transparent way using estimates of the gravity of violence as central elements. The overall approach is portable and frugal in terms of data requirements, which proves advantageous when analyzing conflict zones characterized by economic deprivation and limited data availability. An additional empirical contribution is the documentation and implementation of a procedure to construct bilateral flows of violence from standard place-based microdata on conflicts, such as the Armed Conflict Location and Event Dataset (ACLED).

A benefit of our structural approach compared to standard reduced-form designs is to enable simulations of different counterfactual scenarios. The final part of the paper runs two scenarios and provides novel general equilibrium quantification of their consequences on income, welfare and violence. The structural model also allows for several decomposition exercises, enabling us to quantify the *contribution of each behavioral margin to overall violence*. Our first simulation explores the consequences of a 10% exogenous increase in farming productivity within the Kanuri region, where Boko Haram (one of the most violent groups in Western Africa) operates. This exercise addresses a debated question in the conflict literature regarding how income shocks influence violence. The analysis reveals that competition in the appropriation market shapes the spatial reshuffling of violence, leading to overall pacification of the area that extends well beyond the Kanuri region. Interestingly, productivity in nearby regions also rises as they experience a reduction in violence and

[3]Scholars working on conflicts typically assign the causes of conflict to two categories: greed and grievances (Collier and Hoeffler, 2004). It is now well-documented empirically that greed emerges as a primary driver of violence. The direct form of appropriation encompasses various activities such as looting, extortion, forced labor, and the outright theft of land and resources. This form of appropriation is particularly evident in the control exerted over natural resources like oil and minerals, where competition for resource control becomes a catalyst for conflict (Dube and Vargas, 2013; Berman et al., 2017; Sanchez de la Sierra, 2020). Furthermore, conflicts can be fueled by land disputes arising between transhumant pastoralists and sedentary agriculturalists, which stem from issues related to land tenure and unequal access to fertile land (McGuirk and Nunn, 2024; Eberle et al., 2024; Berman et al., 2021). Numerous other apparent causes of conflicts can also be viewed through the lens of appropriation. Political, cultural, and religious motives are frequently employed as justifications by armed groups; however, they can also serve as a smokescreen to mask acts of appropriation. This dynamic is particularly noticeable in secessionist movements and ethnic conflicts, where the pursuit of resource control and territorial dominance often underlies the surface motivations.

destruction. This spatial diffusion of a productivity shock through reduced violence, an innovative feature of our model, is a first-order and relevant channel for development policy. When the counterfactual accounts for the general equilibrium feedback effects of violence, the predicted welfare increase from the TFP shock in the Kanuri region rises by around 40%.

Our second counterfactual analysis quantifies the expected changes in violence and development due to the "Sahelxit," which occurred on January 28th, 2024, when Burkina Faso, Mali, and Niger withdrew from the CEDEAO (ECOWAS in English) trade agreement. The prospective economic consequences of trade disintegration has been studied extensively recently for the case of Brexit (Sampson, 2017, covers several ex-ante quantifications). However, there has been to our knowledge no evaluation of such policy experiment accounting for the interaction of economic and violence margins, which is important for Sahelxit, given the degree of violence witnessed in the region. Our model addresses this gap by simulating the interactions between trade-induced income changes and violence. We find that exiting the agreement results in a loss of economic opportunities, making looting relatively more attractive and drawing more fighters. Consequently, regions breaking away from CEDEAO experience a significant income drop and exert more violence outwards, while other regions exert less. In net, the level of violence increases in almost all regions in Western Africa, likely further destabilizing this part of the continent.

Literature: The economic literature on conflicts has experienced notable empirical advances over the past fifteen years, driven by the availability of geolocalized data on violent events at a fine-grained level. Among the most recent papers using such data: Dube and Vargas (2013); Michalopoulos and Papaioannou (2016); Berman et al. (2017); Harari and Ferrara (2018); McGuirk and Burke (2020); Moscona et al. (2020); McGuirk and Nunn (2024). These studies typically look at the *local* socio-economic drivers of conflicts and do not explore the spatial determinants of violence.[4] Moreover, they do not explicitly aim to bring formal models to the data, making it difficult to quantify the welfare implications of counterfactual policy interventions.

A small set of papers makes use of quantitative models to analyze conflicts. The pioneering work by Francois et al. (2015) assesses how much coups and revolutions shape cabinet allocations by ethnicity in sub-Saharan Africa. König et al. (2017) present a structural framework that allows to estimate fighting externalities in networks of military alliances and enmities among armed groups in Africa. Adhvaryu et al. (2021) estimate a two-player model of conflict where military capacities are influenced by natural resource endowments. None of these studies model or estimate the impact of spatial frictions. Amarasinghe et al. (2020) and Mueller et al. (2022) explore with a structural approach the theoretical and empirical aspects of violence diffusion in spatial networks of fighting groups. These two papers provide valuable insights into the violent behavior of geographically located fighting groups. However, they adopt a partial equilibrium analysis that captures the impact of economics on conflict in a reduced form, typically represented by an exogenous "prize of the contest". As a consequence, they are silent on the feedback loop from the fighting equilibrium

[4]Some of these papers (Berman et al., 2017, for instance) estimate theory-free models that examine the spatial decay of predictors of violent events, such as the impact of a spike in the global market price of locally produced resources (e.g., minerals or crops) on violence at various distances (e.g., 50km, 100km, 500km).

on the economy. More generally, the quantitative literature on conflicts has not yet explored the various general equilibrium channels through which economic development and violence may potentially interact, despite repeated calls from prominent scholars for a better understanding of these mechanisms (Dell et al., 2014; Burke et al., 2015).[5] Our paper aims at filling this gap by building a quantitative general equilibrium model of violence.

Finally, our paper contributes to the modern literature on spatial economics by extending the existing framework of Quantitative Spatial Models (QSM) to address the context of weakly institutionalized countries. Doing so, we deviate from the standard structure of QSMs (see Redding and Rossi-Hansberg, 2017; Redding, 2022, for recent surveys) by accounting for a key characteristic of these countries, namely the presence of a violent context characterized by inadequate enforcement of property rights and limited state capacity. Our research aligns with the recommendation put forth by Proost and Thisse (2019) in their survey of the spatial economics literature: "when the economy is mainly informal and market institutions do not function well, we need new models."

The remainder of the paper is as follows: Section 2 sets up the theoretical microfoundations of our framework, and shows how the two main gravity equations for trade and violence are derived and also how they interact in the general equilibrium of our model. The gravity equation of bilateral violence is the main empirical novelty of our paper, and is estimated in Section 3, which also describes our data. Section 4 then takes estimated frictions from this estimation and explains how the rest of parameters needed to simulate the model are recovered by combining the structure of the model with observables. Section 5 runs counterfactual simulations. The paper is accompanied by an Online Appendix available on the webpage of the corresponding author.[6]

2 Theory

Our theoretical setup models the interactions between two activities in a multi-regional world: producing tradable goods (farming) and appropriating them (fighting). Individuals choose to allocate their labor to either farming or fighting. Inter-regional frictions hinder both the ability of workers to ship their production to non-local markets, and the ability of fighters to steal income generated at long distance from their home base. These frictions generate two gravity-like equations: one for trade in goods and one for bilateral flows of violent activity. The two equations are interconnected through labor market clearing conditions and optimal occupational choices in each region, ensuring general equilibrium. We begin by describing the gravity equations, and then solve for the equilibrium.

2.1 Gravity of goods

There are N regions indexed by i, with a population of $\bar{L}_i$ that freely allocates between farming (L_i) and fighting (l_i). In this Armington (1969) economy, each region i produces a single variety of the

[5] The seminal contribution by Dal Bó and Dal Bó (2011) considers such general equilibrium interactions in a purely theoretical setup.

[6] The webpage can be found at: https://people.unil.ch/mathiasthoenig/.

farming (tradable) good, and is the sole source of this variety. Consumers in region n have a CES utility which aggregates the quantities q_{in} shipped from producing regions i. It is given by $U_n = \left(\sum_{i=1}^{N}(q_{in})^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$ with $\sigma > 1$. With perfect competition and iceberg trade costs τ_{in}, the market price is equal to $p_{in} = \tau_{in} w_i^P / A_i$, with w_i^P and A_i representing the wage of farmers (producers) and their productivity, respectively. The share of aggregate expenditure E_n that consumers in region n spend on the variety from i is given by:

$$\pi_{in} \equiv \frac{(\tau_{in} w_i^P / A_i)^{1-\sigma}}{\sum_k (\tau_{kn} w_k^P / A_k)^{1-\sigma}}. \tag{1}$$

The equilibrium bilateral trade flow is given by:

$$Y_{in} = \pi_{in} E_n = \tau_{in}^{1-\sigma} \times \left(\frac{w_i^P}{A_i}\right)^{1-\sigma} \times \frac{E_n}{\sum_{k=1}^{N}\left(\frac{\tau_{kn} w_k^P}{A_k}\right)^{1-\sigma}}, \tag{2}$$

which one of the many formulations of the gravity equation for goods belonging to the class of "structural gravity" (Head and Mayer, 2014).[7] The aggregate trade revenues of producing region i (including internal trade) are the sum of trade revenues from all destination regions n:

$$w_i^P L_i = \sum_{n=1}^{N} \pi_{in} E_n, \tag{3}$$

where L_i is the endogenous labor supply of farmers in region i. Together, these equations provide the framework for understanding the partial equilibrium of trade in our model.

2.2 Gravity of violence

Let us now turn to patterns of inter and intra-regional violence, considering violence as an appropriation game between origins i and destinations n. Each origin region potentially hosts a fighting group that recruits l_i fighters at a local competitive wage w_i^F and optimally assigns them to regional "targets" n subject to a spatial friction factor $\xi_{in} > 1$. Each destination region is characterized by its local state capacity, $s_n \in [0, 1]$, which stands for the share of its gross income Y_n that is secured by local authorities through various means, including police, military forces, and property rights enforcement in general. It is assumed to be exogenous in the baseline model.[8] In contrast, the residual share $(1 - s_n)$ of its income is unsecured and is looted by fighting groups which operate in n.

Appropriation is modeled as a repeated game which we briefly outline here (see Appendix A

[7] The most commonly used microfoundations include Anderson and van Wincoop (2003) (which we draw upon here), Krugman (1980), Eaton and Kortum (2002), and Chaney (2008).

[8] In the counterfactual analysis of Section 5.1, we introduce two extensions to the model to account for feedback loops. One extension allows s_n to increase with the level of income Y_n while the other captures the destructive impact of total violence in n on farming productivity A_n.

for more details). Farmers first receive their wages in n (front-loaded payment) and fighters are allocated across regions, initiating a series of repeated contests over the lootable portion of income. Each stage game involves: (i) securing a share s_n of regional income in a *permanent* way; (ii) competition among fighting groups for the remaining unsecured share; (iii) repatriation of fighting revenues to origin regions before the next stage begins. The sequence of contests ends when the residual unsecured income becomes "ε-negligible", resulting in aggregate income being fully secured in all regions.[9] An important distinction in the model is between aggregate gross income Y_n (only partly secured) and aggregate expenditure E_n (fully secured). With the assumed micro-foundations of our model, they are linked by the following relation:

$$E_n = s_n Y_n. \tag{4}$$

Therefore, the aggregate ratio E_n/Y_n is not affected by the endogenous level of fighting. This convenient feature, while not essential, enhances the analytical tractability of the GE framework; alternative micro-foundations are presented in Appendix A.

On the battlefield of region n, the fighting group originating from i produces observable violence with the following constant returns to scale technology:

$$\texttt{violence}_{in} = \psi_i l_{in}, \tag{5}$$

where l_{in} is the number of fighters assigned by i to n and the productivity parameter ψ_i is referred to as i's fighting capacity. In turn, the operational performance of group i on the battlefield is calculated as the ratio of its produced violence to the spatial friction factor ξ_{in}, multiplied by a Frechet unobservable shock, $\tilde{u}_{in}$, with shape parameter $\gamma > 0$ which captures how homogeneous is military luck over different military group-destination combinations. Victory in the battle for lootable income in n goes to the group with largest operational performance. From our functional form assumptions, the probability that i succeeds is:

$$p_{in} \equiv \text{Prob}\left(\frac{\psi_i l_{in}\tilde{u}_{in}}{\xi_{in}} > \frac{\psi_k l_{kn}\tilde{u}_{kn}}{\xi_{kn}}, \quad \forall k \neq i\right) = \frac{(\psi_i l_{in}/\xi_{in})^\gamma}{\sum_{k=1}^{N}(\psi_k l_{kn}/\xi_{kn})^\gamma}. \tag{6}$$

This equation describes the probability of success and hence the incentives for a fighting group in i to launch operations in n. This constitutes the first key element of the gravity of violence. The second brick required to obtain gravity pertains to optimizing the spatial allocation of troops (l_{in}). This allocation aims to maximize i's gross fighting revenues. For tractability, we assume atomistic players. One advantage of this assumption is the simplification of strategic interactions; there is, in particular, no market power on the appropriation market. Each fighting group best-responds to other groups' actions but does not internalize its impact on aggregate violence. Second, the

[9] Dynamically, farmers are first looted by fighters, who are in turn looted by other fighters, and so on. To understand the law of motion of unsecured income, we need to perform an accounting of the looted and repatriated resources across regions at each stage of the game. Note that, for any arbitrary small $\varepsilon > 0$, the end game period T has an order of magnitude given by $\max_n(1-s_n)^T < \varepsilon$. As a result, the model allows for asymptotic convergence only for $\varepsilon = 0$. See Appendix A for details.

continuum assumption implies that the winning probability p_{in} can be interpreted as the realized share of income in n captured by group i.[10] Expected looting revenues of i thus match their realized value, $\mathbb{E}(R_i) = R_i$, and the objective function is given by:

$$R_i \equiv \max_{\{l_{in}\}} \sum_{n=1}^{N} p_{in} \times (1 - s_n) Y_n \qquad \text{s.t.} \qquad l_i = \sum_{n=1}^{N} l_{in}. \tag{7}$$

Looting revenues are then redistributed among fighters. In consequence, we have $R_i = w_i^F l_i$. To obtain a concave maximization problem of R_i with respect to the optimal allocations of fighters l_{in}, we assume $0 < \gamma < 1$, which is supported by our empirical estimate (we find $\hat{\gamma} = 0.453$). This assumption also ensures that an interior solution exists. Solving for the optimal allocation of troops, taking wages as given, and using the resulting l_{in} across regions in (6) yields:[11]

$$\texttt{violence}_{in} \equiv \psi_i l_{in} = \xi_{in}^{-\frac{\gamma}{1-\gamma}} \times \left(\frac{\psi_i}{w_i^F}\right)^{\frac{1}{1-\gamma}} \times \frac{(1 - s_n) Y_n}{\sum_{k=1}^{N} \xi_{kn}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_k}{w_k^F}\right)^{\frac{\gamma}{1-\gamma}}}. \tag{8}$$

This equation for the equilibrium flow of violence from i to n exhibits strong similarities with gravity equations in trade/migration models (the parallel can be seen by comparing with equation 2).[12] However, it is a gravity equation driving the "quantity" of violence, which is what we observe in the dataset at hand. The bilateral violence flows *in value* is $w_i^F l_{in} = \texttt{violence}_{in} \times \left(\frac{w_i^F}{\psi_i}\right)$. This allows to write an expression (parallel to aggregate trade revenues in equation 3) for the aggregate looting revenues of fighters in region i:

$$w_i^F l_i = \sum_{n=1}^{N} \texttt{violence}_{in} \times \left(\frac{w_i^F}{\psi_i}\right). \tag{9}$$

Economic shocks such as changes in wages impact violence in complex ways, channeled by the full spatial structure in equation (8). An increase in income at origin (w_i^F) leads to a decrease in violence emanating from i, while an increase in income at destination (w_n^P, and therefore the amount of appropriable resource Y_n) results in an increase in violence. These two mechanisms, respectively known as the *opportunity cost* and *rapacity* channels, are intuitive and have been extensively tested in the empirical literature (e.g Dube and Vargas, 2013; McGuirk and Burke, 2020). Our spatial framework clarifies that the former pertains to outward violence, projecting from the source, and the latter relates to inward violence, affecting destination. A new effect, captured by the $\sum_k \xi_{kn}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_k}{w_k^F}\right)^{\frac{\gamma}{1-\gamma}}$ term, emerges as a result of competition on the appropriation market among groups from different regions k fighting for income in n. A rise in w_k^F in any region k leads to an increase in violence from i to n through decreased competition on the battlefield (because l_{kn} de-

[10] This modeling strategy is similar to various approaches in trade which micro-found aggregate gravity flows out of individual behavior of heterogeneous agents (see section 2.3 of Head and Mayer, 2014).

[11] The full derivation is provided in Appendix B.

[12] Note that the model allows for all bilateral flows possibilities: local looting whenever $\xi_{ii} < \infty$, raiding whenever $\xi_{in} < \infty$ with $i \neq n$, and even zero violence when $\xi_{in} = \infty$.

creases) and therefore a higher incentives for i to send fighters in n. We refer to the denominator term in equation (8) as *the Multilateral Resistance of Violence*, MRV, in parallel to the terminology used in trade to account for economic competition in a given destination market (Anderson and van Wincoop, 2003, were the first to use this expression). As for gravity in goods regressions, omitting multilateral resistance term (which accounts for spatial interdependence) is a source of misspecification in conflict regressions. This issue is discussed in more detail in the next section.

2.3 General equilibrium

In equilibrium, free occupational choice between farming and fighting leads to equalization of (fully secured) incomes, and thus wages in i:

$$s_i w_i^P = s_i w_i^F = s_i w_i. \tag{10}$$

Gross income can therefore we rewritten as

$$Y_i = w_i^P L_i + w_i^F l_i = w_i \bar{L}_i, \tag{11}$$

with the last equality ensured by the labor market clearing condition under which fighters and farmers add up to total workforce $\bar{L}_i$:

$$\bar{L}_i = L_i + l_i. \tag{12}$$

Two parallel market clearing conditions must also hold in equilibrium. The first one is that the sum of demands for the good produced in i has to add up to the trade revenues of i. This is equation (3), which—combined with equations (2), (4), (10), and (11)—can be rewritten as

$$w_i L_i = \sum_{n=1}^{N} \frac{\tau_{in}^{-(\sigma-1)} \left(\frac{A_i}{w_i}\right)^{\sigma-1}}{\sum_{k=1}^{N} \tau_{kn}^{-(\sigma-1)} \left(\frac{A_k}{w_k}\right)^{\sigma-1}} s_n w_n \bar{L}_n. \tag{13}$$

The second market clearing condition is that the bilateral flows of revenues obtained from appropriation sum to total revenues of violence landing in i. Combining (8), (10), and (11), and inserting the result into (9), we obtain:

$$w_i l_i = \sum_{n=1}^{N} w_i l_{in} = \sum_{n=1}^{N} \frac{\xi_{in}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_i}{w_i}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_{k=1}^{N} \xi_{kn}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_k}{w_k}\right)^{\frac{\gamma}{1-\gamma}}} (1 - s_n) w_n \bar{L}_n. \tag{14}$$

The general equilibrium is characterized by the system of $3 \times N$ equations (12)–(13)–(14) and $3 \times N$ unknowns $\{L_i, l_i, w_i\}$, for $i = 1, \ldots, N$.

Existence and uniqueness: We can characterize the general equilibrium more compactly using a fixed-point "master equation" system with the wage vector as the sole endogenous variable. To achieve this, we combine the fighting and trade revenue equations by summing equations (13) and

(14), yielding

$$w_i \bar{L}_i = \sum_{n=1}^{N} \beta_{in}(\mathbf{w}) w_n \bar{L}_n, \tag{15}$$

where $\beta_{in}(.)$ are non-linear functions of the wage vector $\mathbf{w}$:

$$\beta_{in}(\mathbf{w}) \equiv (1 - s_n) \times \frac{w_i^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_i}{\zeta_{in}}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_{k=1}^{N} \left(\frac{\psi_k}{\zeta_{kn} w_k}\right)^{\frac{\gamma}{1-\gamma}}} + s_n \times \frac{w_i^{1-\sigma} \left(\frac{A_i}{\tau_{in}}\right)^{\sigma-1}}{\sum_{k=1}^{N} \left(\frac{A_k}{\tau_{kn} w_k}\right)^{\sigma-1}}. \tag{16}$$

To prove existence and uniqueness of a general equilibrium in the wage vector, we rely on resolution techniques from spatial economics. Specifically, in Appendix C, we use the isomorphism between the trade and fighting revenue equations in order to apply Mas-Colell et al. (1995) and Alvarez and Lucas (2007). We conclude that a general equilibrium exists and is unique in this baseline version of the model.

Monadic violence: Equation (8) describes the bilateral flow of violence, i.e., the equilibrium number of fighters from i allocated to appropriate income in n (l_{in}), adjusted for their fighting capacity (ψ_i). In many research designs, the only outcome observable to the researcher is the total (monadic) violence exerted in region n, $\texttt{violence}_n = \sum_{i=1}^{N} \texttt{violence}_{in}$. Our bilateral framework can also be useful to guide the analysis of violence determinants in those cases when the origin is not observable. To see this, note that with free mobility of individuals between farming and fighting, equation (8) can be combined with equations (10) and (11) and rewritten as:

$$\texttt{violence}_{in} = \omega_{in} \times \left(\frac{\psi_i}{w_i}\right) \times (1 - s_n) w_n \bar{L}_n, \qquad \text{with} \qquad \omega_{in} \equiv \frac{\zeta_{in}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_i}{w_i}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_{k=1}^{N} \zeta_{kn}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_k}{w_k}\right)^{\frac{\gamma}{1-\gamma}}}, \tag{17}$$

which, once summed over origins and logged, gives

$$\log\left(\texttt{violence}_n\right) = \log(1 - s_n) + \log \bar{L}_n + \log w_n + \log\left(\sum_i \omega_{in} \frac{\psi_i}{w_i}\right). \tag{18}$$

The first three terms in equation (18) are intuitive: violence falls with the capacity of local authorities to secure income (s_n), and rises with the local aggregate income (log of $w_n \bar{L}_n$), which attracts groups to fight over its non-secured part (the *rapacity* effect). Consequently, logged per capita income $\log w_n$ naturally enters the equation with a unitary coefficient. The last term of equation (18) represents a "Supply Potential of Violence" (SPV), equal to the weighted average of the fighting potentials of all origins i (including n itself) attempting to appropriate income in n (the spatial weights ω_{in} are defined in equation 17). Income per capita w_i enters SPV negatively, as it represents the *opportunity cost* of being a fighter versus being a farmer in the origin regions. Intuitively, the proximity of poor regions with strong fighting capacity (high ψ_i / w_i), indicated by a low ζ_{in} in

ω_{in}, contributes to raising violence in n.[13]

Equation (18) suggests that empirical work intending to estimate the impact of local income, w_n, on violence occurring in n should incorporate the SPV term (or at least proxy for it). The first reason is that the SPV includes the local opportunity cost of becoming a fighter in region n. Secondly, the SPV term captures (within our model) *all* spatial interactions that might act as confounders when trying to tease out the impact of local income on local violence. In our empirical analysis (Section 3), gravity estimates reveal that physical distance is an important component of the frictions to bilateral violence. Therefore, a simple geographic proxy for the SPV to include into a monadic regression of violence could be

$$\widehat{\texttt{SPV}}_n \equiv \log\left(\sum_{i=1}^{N} \frac{\texttt{dist}_{in}^{\hat{\alpha}}}{\sum_{k=1}^{N}\texttt{dist}_{kn}^{\hat{\alpha}}} w_i^{-1}\right),$$

where $\hat{\alpha} < 0$ is the estimated impact of distance on bilateral violence, which we find to be close to -2.5 in Table 2.

3 Empirical Gravity of Violence

3.1 Data sources and construction

Our data construction procedure is made of six steps, which are outlined in Table 1. Each of these steps is briefly presented below (steps 1-4) and in the next section (steps 5-6), with Online Appendix Section OA1 providing all the details on the data and procedure.

Conflict data: We use the Armed Conflict Location and Event Dataset (ACLED) which covers conflict events in all African countries from 1997 to June 2023.[14] This dataset is extensively used in academic work (see footnote 1). We innovate by constructing bilateral flows of violence with well-defined origins and destinations. Essentially, we use the coordinates of the violent events to determine the destinations, while the identities of the actors who perpetrate this violence, combined with external information sources, help identify the origins. ACLED reports information about the date, GPS location, type of violence, fatalities, and actors involved in each violent event. It is compiled from various primary sources, including press accounts from regional and local news, humanitarian agencies, and research publications. Following standard practices, our quantitative analysis uses the number of events as the metric of violence, rather than fatalities, which

[13] When violence is fully local, the *rapacity* and *opportunity cost* effects exactly compensate each other and income has no effect on violence. Algebraically, this limit case corresponds to $\xi_{nn} > 0$ and $\xi_{in} = \infty$, $\forall i \neq n$. There, the spatial weights load entirely on the local component $\omega_{nn} = 1$ (zero when $i \neq n$) and equation (18) becomes $\log(\texttt{violence}_n) = \log(1 - s_n) + \log \bar{L}_n + \log \psi_n$. More generally, when non-local frictions are significantly larger than local ones ($\xi_{in} \gg \xi_{nn}$), the *rapacity* and local *opportunity cost* effects tend to offset each other. As a consequence, the impact of local income per capita, w_n, on violence is lowered. Those countervailing forces are determined by the general equilibrium of the model, which speaks to the ambiguity highlighted by the quote of footnote 2.

[14] The dataset "*Compatibility*" has been downloaded from https://acleddata.com/ on June 28, 2023 (2pm).

are known to be imprecisely recorded at a disaggregated level.[15] Nevertheless, for completeness, we also present some aggregate statistics on fatalities in this section. ACLED captures a broad spectrum of violence, ranging from anti-government protests to high-intensity battles or civilian attacks by paramilitary groups. The type of violent events and actors thus vary *significantly* within the dataset. In consequence, beyond cleaning and preparing data for our spatial analysis (steps 1-3), our work also consists of isolating the type of violence that specifically relates to the spatial projection of armed groups (steps 4-6).

Table 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...).

The raw data (step 0) feature 325541 distinct violent events involving a large and heterogeneous set of 14137 distinct actors.[16] ACLED considers three levels of precision for the spatial location of events. We keep only those with the highest level of precision, i.e., town level, which amount to roughly three-quarters of the total. We also exclude events that are not located in continental Africa. After applying these two *geographic filters* (step 1), we end-up with 237489 events covering 11138 actors.

ACLED reports information on the nature of violence attached to each event.[17] However, a small fraction of events are in fact categorized as non-violent or non-identified (respectively "Agreement" and "Others" in the ACLED nomenclature). The *events filter* (step 2) excludes those events. After this cut, the sample comprises 234943 events involving 11013 actors. In step 3, we remove violent actors that are not assigned by ACLED to a precise politico-military category (see

[15] The ACLED Codebook (2023, p.37) accordingly advises that information on fatalities should be considered with caution.

[16] On average, each event involves 2.3 actors; but this number can rise substantially for some major battles. ACLED reports information on 8 categories of actors: State Forces, Rebel groups, Political Militias, Identity Militias, Rioters, Protesters, Civilians, External/Other Forces. See Table OA1.2 and footnote 2 in the Online Appendix for more details.

[17] ACLED records six types of violent events (and 25 sub-event types): Battles, explosions/remote violence, protests, riots, strategic developments and violence against civilians. See footnote 3 in online appendix for more details.

footnote 16). This *information filter* takes out many actors but relatively few events because, in a typical ACLED event, at least one actor belongs to a well-identified category. We retain 232878 distinct events covering 5486 actors.

In line with our theoretical model of spatial 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. 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.[18] 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.

Other datasets: Information on nighttime light data is obtained from the harmonized global nighttime light dataset (Li and Zhou, 2017). Population data is obtained from WorldPop using the top-down unconstrained estimation modeling approach for 2000 (Lloyd et al., 2019). We identify the main crop(s) produced by each cell using data from the FAO's Global Agro-Ecological Zones (Fischer et al., 2008). This dataset is constructed from models that use location, climate information, and soil characteristics to generate a global GIS raster of cell-level suitability for 17 crops for which we have worldwide prices. Crop prices (base 100 in 2000) are obtained from the World Bank Commodity Dataset (World Bank Group, 2020).[19]

3.2 Vectorization of georeferenced conflict events

Here we describe our procedure of "vectorization" that processes ACLED violent events into bilateral flows of violence (denoted $\texttt{violence}_{in}$) with a magnitude and a direction (from i to n). In the model, each fighting group g is associated with a rear base $rb(g) \in i$, where recruitment takes place and fighters are paid the prevailing competitive wage of the local labor market in the region of origin i. The bilateral flow of violence between i and n is defined as

$$\texttt{violence}_{in} = \sum_{g} \mathbb{I}_{rb(g)\in i} \quad \times \quad \sum_{t=1997}^{2023} \#\texttt{events}_{gnt}, \tag{19}$$

where $\#\texttt{events}_{gnt}$ is the number of events attributed to group g, perpetrated in n over time period t, and $\mathbb{I}_{rb(g)\in i}$ is an indicator which takes a value of 1 if the rear base of the group g is located within region i. This defines a cross-section of flows of violence between geographical units i and n as our dependent variable.

The theory suggests that $rb(g)$ should be the location of the labor pool that g uses to hire fighters. In the absence of information on the market for fighters at the scale of the African continent,

[18] 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.

[19] Night time light data: https://www.mdpi.com/2072-4292/9/6/637; World Pop: https://hub.worldpop.org/geodata/listing?id=64; FAO's Global Agro-Ecological Zones: https://gaez.fao.org/

we assume that an armed group's primary basin of recruitment stems from its ties to ethnic groups. This view is widely held in the literature and supported by extensive empirical evidence.[20] There are several reasons explaining why the recruitment of armed groups follows ethnic lines. Like any labor relationship, leaders typically recruit from their existing networks, often grounded in ethnicity in Africa. This enhances the enforceability of contracts (Caselli and Coleman, 2013) and leads to more committed rebels (Weinstein, 2005, 2007). Moreover, social incentives and ethnographic structure within ethnic communities contribute to intrinsically motivate individuals to participate in conflicts (Sambanis, 2001; Moscona et al., 2020).

In practical terms, we first map armed groups g to ethnic groups with the help of external informational sources.[21] Then, we assign rear bases $rb(g)$ to historical ethnic homelands as observed on Murdock's map (Murdock, 1959). It is important to note that our approach does not presuppose that violence necessarily arises from ethnic cleavages. It rather assumes that the spatial scope of fighters' recruitment is constrained by the borders separating ethnic groups. We have undertaken substantial data work to identify the ethnic affiliation of armed groups—one empirical contribution of this paper being to make this mapping available to the research community. Our work was greatly facilitated by the granularity of the violence data, a feature not previously documented in the literature. This point is illustrated in Figure 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.[22] Limiting the collection of ethnic ties to this set of major actors is therefore not very costly in terms of the coverage of actual violent events.

The restriction of the construction of rear bases (and therefore origins of violence) to this set of 220 actors leads to a new sample cut, referred to as the *Murdock filter on actors* (step 5) in Table 1. We find that these 220 perpetrators of violence are associated to 87 different origins (Murdock ethnic regions). In order to identify destinations of violence, we must project violent events onto

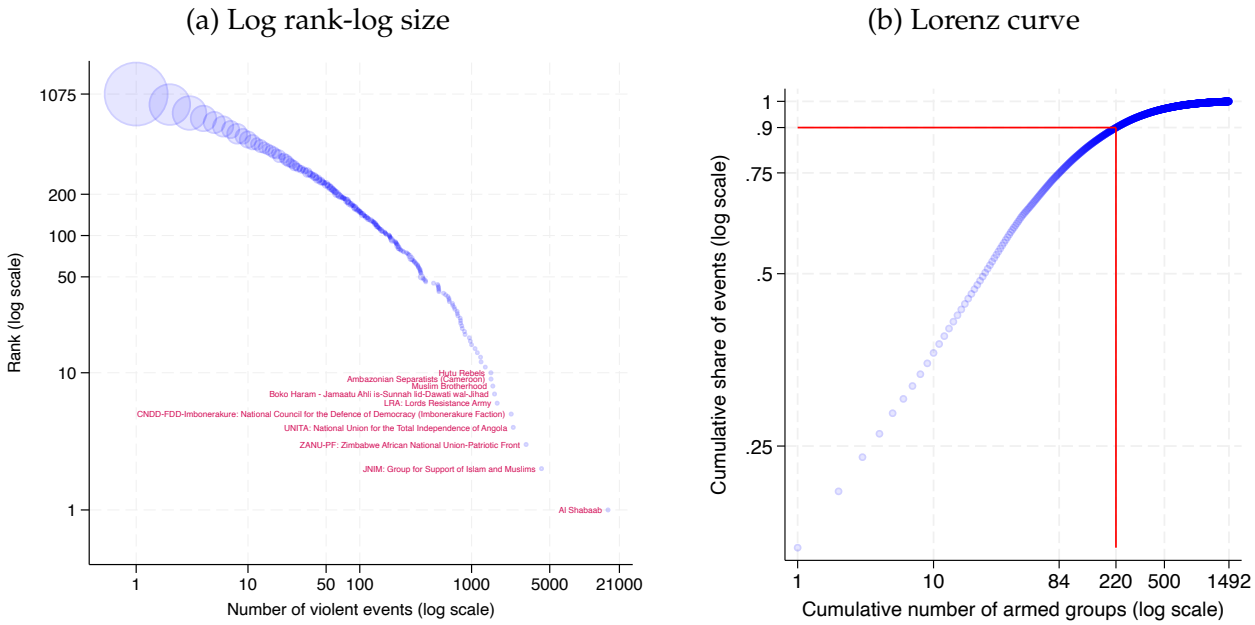
[20] A substantial body of literature underscores the importance of mobilizing fighters among co-ethnics (Caselli and Coleman, 2013; Michalopoulos and Papaioannou, 2016; Mueller et al., 2022; Guarnieri and Tur-Prats, 2023). Additionally, several empirical studies document how ethnic mobilization serves as a recruitment method in Africa (Fearon and Laitin, 1999; Sambanis, 2001; Elbadawi and Sambanis, 2000; Gates, 2002, 2017; Cederman et al., 2010). The Online Appendix Section OA2 presents supportive anecdotal evidence.

[21] The conventional approach in the recent literature interested in the ethnic affiliations of violent actors typically involves matching the violent events perpetrated by each actor with maps delineating ethnic regions (Michalopoulos and Papaioannou, 2016; Moscona et al., 2020; Adhvaryu et al., 2021; Mueller et al., 2022; Eberle et al., 2024; McGuirk and Nunn, 2024; Gehring and Schaudt, 2024). In contrast, our approach defines the ethnic ties of actors through external sources that are *independent* of the violence observed and recorded in ACLED. This ensures that these ethnic ties are not influenced by any contamination from the observed violence, which is the focus of our structural model analysis. In this regard, we follow the methodology of Berman et al. (2017), who identified ties to ethnic groups for 109 violent actors active between 1997 and 2010, primarily relying on ACLED information and the ethnicity of the group's leaders and troops. Using a broader range of informational sources, we manage to cover 90% of the violence in Africa perpetrated by rebel groups and political militias. Similarly, Sestito (2024) also relies on external sources to recover the ethnic ties of non-state armed groups active in a subsample of 18 African countries between 2005 and 2015.

[22] The step 4 sample comprises 1492 violent actors who are involved in 87578 events resulting in 379314 fatalities. Among them, the 220 most violent actors participate in 80619 events leading to 362672 fatalities (92% of the events, 96% of the reported fatalities).

ethnic homelands. Consequently, we remove from the sample all events that do not intersect with the Murdock map (e.g., Sinai region, violence occurring in lacustral areas). After this *Murdock filter on events* (step 6), we end up with 78335 distinct events. Note that almost 80% of this violence is directly exerted against civilians or takes the form of military battles (Table OA1.3). Based on this sample, the final step of the vectorization procedure involves applying equation (19) to construct bilateral flows $\texttt{violence}_{in}$. This variable represents the total number of violent events recorded between 1997 and June 2023 for each origin-destination pair in the set of 824 ethnic regions covered by the Murdock's map. Out of this universe of flows, 97.46% are zero-violence flows, which is not surprising given that many pairs are separated by very large distances (the straight-line distance between Cairo, Egypt, and Cape Town, South Africa, is about 7800 kilometers). Restricting to the subsample of non-zero flows leaves us with 87 origins of violence and 623 destinations of violence (mapped in Figure 2).

Figure 1: The distribution of violent events



Note: Both panels are based on the sample of step 4 in Table 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.

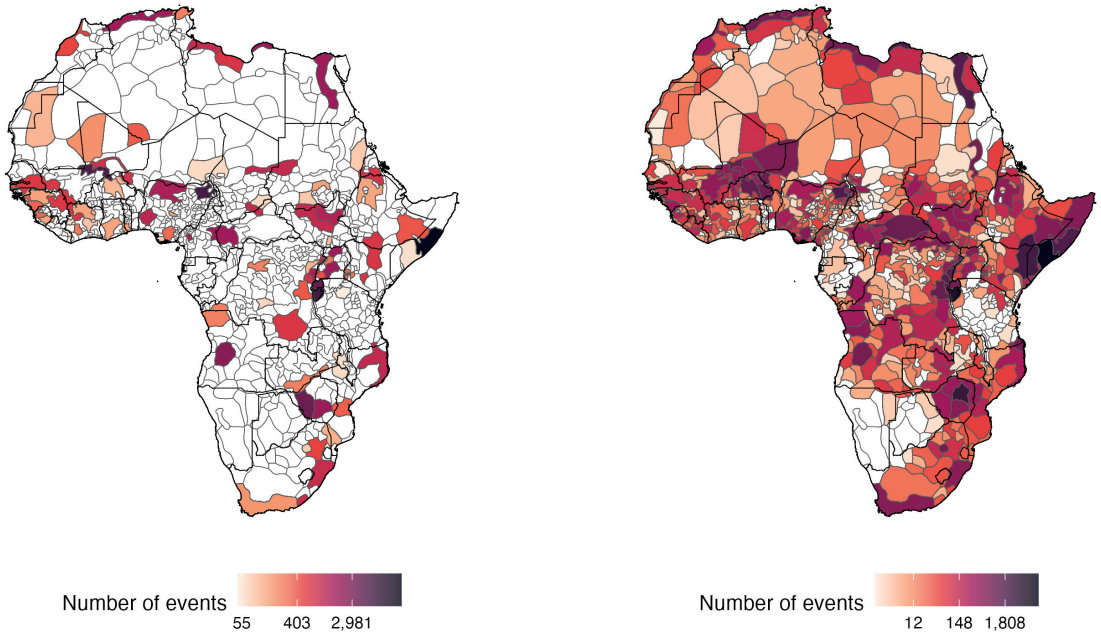
3.3 Econometric equation

As a first motivation for our specification of frictions, Figure 3 presents evidence of unconditional correlation between $\log \texttt{distance}_{in}$ and the bilateral share of events at destination $\frac{\texttt{violence}_{in}}{\texttt{violence}_{n}}$ (our dependent variable in the regressions). In the overall sample, the binscatter in panel (a) shows a strong negative correlation (explaining 80% of the variance of the share of events), which is main-

Figure 2: Origins and destinations of violent events

(a) 87 Origins of violence

(b) 623 Destinations of violence

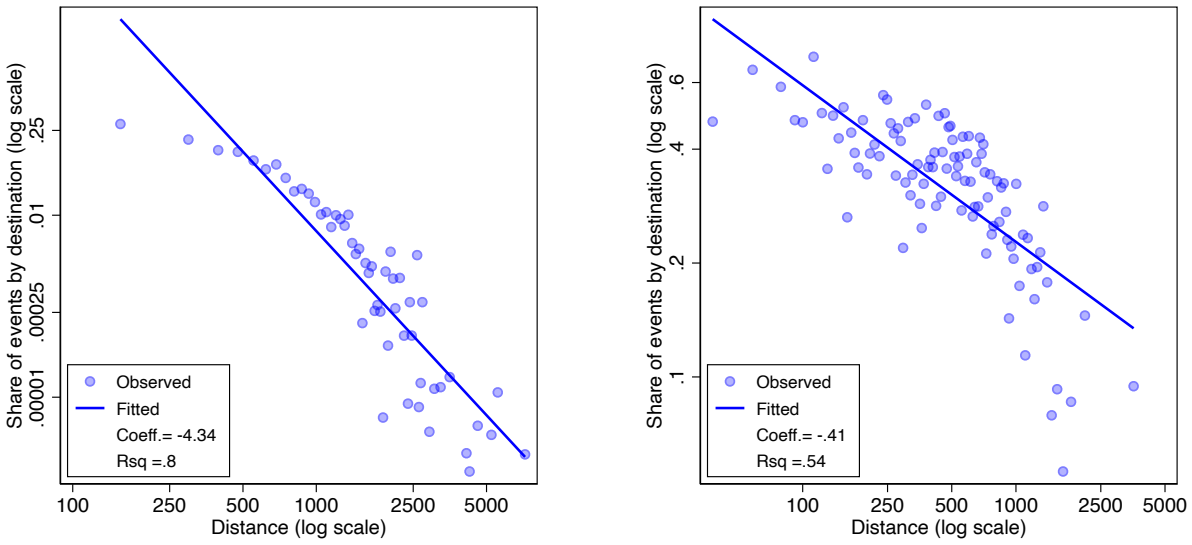


Note: These maps are based on the sample of step 6 in Table 1. It comprises a total of 78335 violent events originating from 87 regions and targeting 623 destinations.

Figure 3: Bilateral distance and share of events at destination

(a) Full sample

(b) Sample with shares > 0



Note: These figures are based on the sample of step 6 in Table 1. It comprises a total of 78335 violent events originating from 87 regions and targeting 623 destinations. Both panels display binscatter with 50 bins.

tained in panel (b) when we restrict the sample to strictly positive flows. Our econometric specification will allow for this friction and add the possibility for ethnic and national borders to also affect bilateral violence. We now turn to describing our estimating equation.

Based on our theoretical prediction, we estimate a gravity equation to reveal the empirical determinants of bilateral violence. Building on equation (8), we obtain:

$$\frac{\texttt{violence}_{in}}{\texttt{violence}_{n}} = \frac{\psi_i l_{in}}{\sum_{k=1}^{N} \psi_k l_{kn}} = \xi_{in}^{-\frac{\gamma}{1-\gamma}} \times \left(\frac{\psi_i}{w_i}\right)^{\frac{1}{1-\gamma}} \times \left[\sum_{k=1}^{N} \xi_{kn}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_k}{w_k}\right)^{\frac{1}{1-\gamma}}\right]^{-1}. \tag{20}$$

We follow the now standard practice in the gravity literature to estimate this equation using Poisson pseudo-maximum likelihood (PPML) with high-dimensional fixed effects,[23] translating our estimating equation into:

$$\mathbb{E}\left(\frac{\texttt{violence}_{in}}{\texttt{violence}_{n}}\right) = \exp\left[-\frac{\gamma}{1-\gamma}\log \xi_{in} + \texttt{FE}_i^o + \texttt{FE}_n^d\right], \tag{21}$$

where the dependent variable is the share of violence in n which comes from i. The regression provides estimates of spatial frictions ξ_{in} together with origin and destination fixed effects ($\texttt{FE}_i^o$ and $\texttt{FE}_n^d$), which are crucial objects for the next step when we turn to the task of recovering the structural parameters of the model.[24] We consider a specification where ξ_{in} includes three observable frictions: bilateral distance, ethnic homeland border-crossing, and country border-crossing.[25] This leads to our empirical gravity equation of violence:

$$\begin{aligned}\mathbb{E}\left(\frac{\texttt{violence}_{in}}{\texttt{violence}_{n}}\right) = \exp\Big[&-\frac{\alpha_1\gamma}{1-\gamma}\log \text{distance}_{in} - \frac{\alpha_2\gamma}{1-\gamma}\text{Ethnic border}_{in} - \frac{\alpha_3\gamma}{1-\gamma}\text{Political border}_{in} \\ &+ \texttt{FE}_i^o + \texttt{FE}_n^d\Big],\end{aligned} \tag{22}$$

where the independent variables are the log of the bilateral distance between i and n, a binary variable that equals one if i and n are different ethnic groups, and another binary variable that equals one if i and n are in different countries.

3.4 Results

Table 2 displays the estimation results of equation (22). The sample consists of the cross-section of 87 origins and 623 destinations, for a total of 54201 dyads.[26] Estimation is carried out using the HDFE PPML estimator. Standard errors are clustered by dyad. The elasticity of violence

[23] Santos Silva and Tenreyro (2006) first recommended that estimator for gravity equations based on its robustness to non-homoskedastic error terms. Fally (2015) later showed PPML to be a natural estimator for micro-founded structural models of gravity, allowing for a structural interpretation of the fixed effects as well as large shares of zeroes.

[24] The structural interpretation of those fixed effects is $\texttt{FE}_i^o = \frac{1}{1-\gamma}\log\left(\frac{\psi_i}{w_i}\right)$, and $\texttt{FE}_n^d = -\log\left[\sum_{k=1}^{N} \xi_{kn}^{-\frac{\gamma}{1-\gamma}}\left(\frac{\psi_k}{w_k}\right)^{\frac{1}{1-\gamma}}\right]$.

[25] In Murdock's map, many ethnic groups are split across different countries (Michalopoulos and Papaioannou, 2016). To build the variable of border-crossing transparently, we opt for a simple coding rule: we assign each ethnic group to the country where it predominantly belongs in terms of area. We leave for future analysis the estimation of the heterogeneous impact of border-crossing when ethnic groups are more or less split across countries.

[26] The inclusion of fixed effects leads to dropping from the sample all regions that exert *and* experience no violence.

Table 2: Estimates of the frictions to violence

Model	(1)	(2)	(3)	(4)
log distance	-2.65***	-2.90***	-2.42***	-2.49***
	(0.046)	(0.052)	(0.059)	(0.080)
Ethnic border		2.46***	2.52***	
		(0.181)	(0.176)	
Pol. border			-1.69***	
			(0.109)	
Ethnic border: contiguity				2.06***
				(0.185)
Ethnic border: no contiguity				2.81***
				(0.207)
Political border: contiguity				-1.25***
				(0.096)
Political border: no contiguity				-3.08***
				(0.194)
Observations	54,201	54,201	54,201	54,201
Pseudo R^2	0.517	0.526	0.546	0.555

Note: Dependent variable is the share of events in n originating from i. Cross-section over 1997-2023. Estimated with PPML with i and n fixed effects. Columns (1)-(4) are estimated on the sample of 87 origins attacking a total of 623 destinations. Significant at 10%, ** significant at 5%, *** significant at 1%. Standard errors clustered at the country pair level.

with respect to distance is estimated to be -2.65 (column 1). This implies that doubling the distance between two ethnic groups is associated with a 84% drop in the bilateral flow of violence ($\exp(-2.65 \times \log 2) = 0.16$).[27] To put this elasticity in context, it is useful to compare it to the estimated impact of distance on trade flows. Head and Mayer (2014) report that the elasticity of trade to distance (not distinguishing by transport mode) is -1.1 on average in a meta-analysis of 328 estimates. The elasticity of trade using land transport (mostly trucks in advanced economies) is estimated to be around -2 by Combes et al. (2005). This comparison suggests that the logistics of violence in Africa, which heavily relies on road transportation, may be a significant factor driving the observed spatial patterns of violence.

From columns (2) to (4), we sequentially add variables capturing the frictions associated with crossing an ethnic or a political border. Crossing ethnic borders increases the likelihood of violence by roughly an order of magnitude ($\exp(2.46) = 11.70$), which suggests that raiding other ethnic groups is a key driver of violence. On the other hand, crossing political borders leads on average to a five-fold drop in violence ($\exp(-1.69) = 0.18$), indicating that national border security measures can be effective in reducing violence. In column 4, we breakdown these two border-crossing frictions by distinguishing between contiguous and non-contiguous ethnic groups and

[27] This result aligns with Mueller et al. (2022)'s findings regarding the importance of distance in the geography of violence. However, their approach differs significantly: they estimate the spatial decay of violence as a function of distance to ethnic borders using a monadic regression of violence. These methodological differences limit the possibility of a quantitative comparison between their results and our gravity coefficient of distance.

countries. Moving across ethnic borders towards a non-contiguous ethnic group significantly increases the flow of violence more than moving towards a contiguous ethnic group ($\exp(2.81) = 16.61$ vs $\exp(2.06) = 7.85$). One plausible interpretation for this difference is that contiguous ethnic groups often share common history, culture, or economic interests (e.g. trade), which reduces the incentive to export violence. In contrast, non-contiguous ethnic groups are less interconnected culturally and economically, potentially heightening tensions. Crossing political borders towards a non-contiguous country reduces the flow of violence much more than moving towards a contiguous country ($\exp(-3.08) = 0.05$ vs $\exp(-1.25) = 0.29$). Here, a natural explanation for this pattern in the data is that the number of borders to be crossed through *land transportation* is mechanically larger when countries are non-contiguous. Overall, these findings underscore the importance of considering both ethnic and political factors in understanding violence. They also suggest that policies aimed at securing borders and transport infrastructure could effectively disrupt raiding logistics and reduce violence in conflict-affected regions.

4 Model Inversion

This section presents the procedure for recovering all the structural parameters of the model needed for the counterfactual analysis in the next section. The sample comprises 824 regions, of which 623 experience violence originating from 87 of them. The procedure involves six steps that we detail below. Additional material is relegated to Appendix D, with Table D3 summarizing the results.

Step 1: We estimate the gravity equation of violence (22) to obtain the elasticity of each spatial friction and the set of origin fixed effects $\widehat{\text{FE}}_i^o$ (column 3 of Table 2). For the fixed effects to be comparable, origins of violence in the gravity sample must belong to the "largest connected set", which is constructed using connections formed by multiple origins attacking a given destination (the equivalent of mobility of workers between firms in the labor literature started with Abowd et al., 1999). We have verified that this is the case for all the 87 origins of violence. By construction, one origin of violence, hereafter indexed by $i = 1$, must be taken as a reference with its fixed effect set to zero in the gravity estimation: $\widehat{\text{FE}}_1^o = 0$. It is important to select an ethnic group for which we can retrieve reliable external information regarding the total number of fighters associated with it (see step 2). Therefore, we choose as reference the Zaghawa ethnic group, which is located in both Sudan and Chad. In Figure D1, we plot the estimated $\widehat{\text{FE}}_i^o$ against the observed (logged) number of violent events originating from i. As expected, their correlation is high and positive, confirming that $\widehat{\text{FE}}_i^o$ captures the average propensity of each origin to perpetrate violence.

Step 2: We then recover the parameters γ and ψ_i by using the theoretical correspondence of $\widehat{\text{FE}}_i^o$ obtained from equations (20) and (21):

$$\widehat{\text{FE}}_i^o - \widehat{\text{FE}}_1^o = -\frac{1}{1-\gamma}\log\left(\frac{w_i}{w_1}\right) + \frac{1}{1-\gamma}\log\left(\frac{\psi_i}{\psi_1}\right), \tag{23}$$

where $\widehat{\text{FE}}_1^o$ is reported for completeness but is in practice taken as the reference. This theoretical relation suggests regressing the FEs on local wages to obtain the parameters of interest. However, wages are not observable at the fine geographical scale of our analysis for the whole African continent. Instead, we use nighttime lights per capita as a proxy, assuming $\frac{w_i}{w_1} = \left(\frac{\texttt{light}_i}{\texttt{light}_1}\right)^{\lambda}$. Plugged into (23), this leads to the following empirical equation:

$$\widehat{\text{FE}}_i^o = -\frac{\lambda}{1-\gamma} \log\left(\frac{\texttt{light}_i}{\texttt{light}_1}\right) + \text{residual}_i. \tag{24}$$

We estimate this equation via two-stage least squares (2SLS) to mitigate any reverse causation effects from violence (FE) to development (light). The instrumental variable for $\texttt{light}_i$ is the (log of) average world price of the most suitable crops:

$$\log\left(\overline{\text{Price}_i}\right) = \log\left(\sum_{t=1997}^{2023}\sum_{c=1}^{5} \alpha_i^c \times P_t^c\right),$$

where crop c is one of the five most suitable crops produced in ethnic homeland i, α_i^c is its relative suitability (as measured by FAO data) and P_t^c is its worldwide real price, base 100 in 2000.

Table D1 in Appendix D reports the estimation results of equation (24) for OLS in column (1), first stage in column (2), reduced form version of the regression in column (3), and 2SLS in column (4). The 2SLS coefficient is equal to -0.493, which corresponds to $-\frac{\lambda}{1-\gamma}$ in the theory. The parameter λ is calibrated based on Bruederle and Hodler (2018) who estimate the correlation between wages and nighttime light to be 0.27. Our estimate of γ is therefore $\hat{\gamma} = 0.453$, which is consistent with negative effects of violence frictions on violence flows (a minimal result for the model not to be rejected by the data, considering the exponent on ξ_{in} in equation 20). Finally, we use the residuals of the 2SLS regression to recover the values of $\hat{\psi}_i$. Comparing equations (23) and (24) shows that $\text{residual}_i = \frac{1}{1-\gamma}\log\left(\frac{\psi_i}{\psi_1}\right)$. Note that, to allow for this structural interpretation of the residual, the constant term is not included in the specifications reported in Table D1. The fighting capacity of the reference group, ψ_1, is obtained from external sources.[28] This number, together with the residual and $\hat{\gamma}$, are plugged into the previous formula to get $\hat{\psi}_i$ for each of the 87 origins of violence. For the other regions, we set $\hat{\psi}_i = 0$ as they produce no violence. The sample mean of $\hat{\psi}_i$ for violent regions is equal to 0.014; according to equation (5), our model interprets the inverse of this value, which is approximately equal to 71, as the average number of fighters required to produce one ACLED

[28] ACLED gives a total of 944 violent events originating from the Zaghawa (our reference group). Our procedure assigns the following violent actors to the Zaghawa: (i) Justice and Equality Movement (35000 fighters); (ii) SLM/A: Sudan Liberation Movement/Army and the two factions that resulted from its split in the early 2010s, the Abdul Wahid al-Nur Faction and the Minnawi Faction (5000 fighters all together); (iii) SRF: Sudan Revolutionary Front (60000 fighters); (iv) the political coalition Forces for Freedom and Change (FFC). From our sources, the FFC, established in 2019 to coordinate calls for the removal of President Omar al-Bashir from power and to negotiate a power-sharing plan, does not have its own military force. In line with the theoretical definition (5), the parameter $\hat{\psi}_1$ is computed as the violence-to-soldiers ratio of the Zaghawa group: $\hat{\psi}_1 = 944/100000 = 0.00944$. We use the following sources for collecting information on the number of fighters: https://www.aljazeera.com/news/2010/5/15/who-are-sudans-jem-rebels
https://ucdp.uu.se/additionalinfo/469/0
https://en.wikipedia.org/wiki/Sudan_Revolutionary_Front
https://en.wikipedia.org/wiki/Forces_of_Freedom_and_Change

event.

Step 3: We compute the unsecured share of income based on the fact that the total fighting revenues across all the fighters active in n sum up to the lost income of n:

$$\sum_{i=1}^{N} w_i l_{in} = (1 - s_n) w_n \bar{L}_n. \tag{25}$$

We isolate s_n in equation (25), and use the relationship between wages and nightlights and the theoretical definition of violence (equation 5) to recover an estimate of the state capacity as:

$$\hat{s}_n = 1 - \frac{\sum_{i=1}^{N} (\texttt{light}_i)^{\lambda} \frac{\texttt{violence}_{in}}{\hat{\psi}_i}}{(\texttt{light}_n)^{\lambda} \texttt{pop}_n}, \tag{26}$$

where $\bar{L}_n$ is approximated by total population $\texttt{pop}_n$ from the World Pop dataset (Lloyd et al., 2019), ensuring that all elements of the right-hand-side of the equation are estimated, calibrated or observed. Note that the previous relation implies $\hat{s}_n = 1$ for regions which experienced no violence. For the others, Table D3 shows that their $\hat{s}_n$ falls within the (0,1) range, which is reassuring for the overall consistency of the procedure. The minimal state capacity is associated with the Bira ethnic group, only able to secure 62.8% of their total income.

Step 4: Equations (5) and (12) allow to compute the number of fighters and farmers in each region:

$$\widehat{l}_i = \frac{\sum_{i=1}^{N} \texttt{violence}_{in}}{\hat{\psi}_i} \quad \text{and} \quad \widehat{L}_i = \texttt{pop}_i - \widehat{l}_i, \tag{27}$$

where we set $\widehat{l}_i = 0$ for regions that exert no violence. To gain insights, we report in Table D3 some statistics related to the share of fighters in the population, computed as $\widehat{l}_i / \texttt{pop}_i$, for the 87 origins of violence. With a mean and standard deviation equal to 0.078 and 0.159, these estimates appear to be quantitatively realistic (an average share of fighters of 8% conditional on exerting violence). Importantly, they all fall within the (0,1) range, providing a reassuring sanity check.

Step 5: This step focuses on calibrating the trade module of the model. Inspection of the theoretical equations outlined in Section 2.3 reveals that the GE characterization relies on trade frictions (the τs). In line with common practice in the gravity literature (see Head and Mayer, 2014, for a review of estimates of those frictions), we specify the log of these frictions as the sum of three standard trade impediments:

$$(1 - \sigma) \log \hat{\tau}_{in} = \mu_1 \times \log \text{distance}_{in} + \mu_2 \times \text{Political Border}_{in} + \mu_3 \times \text{RTA}_{in}, \tag{28}$$

where the first two variables on the right-hand side are defined in equation (22); they act as impediments to trade. The third variable, a trade facilitator, is a binary variable that equals one if regions

i and n are located in different countries that are members of the same regional trade agreement; note that this RTA effect will be the one we shock later in one of our counterfactual simulations.

Since we do not observe bilateral trade flows at the ethnic region level, we have to rely on national flows to calibrate the parameters (μs) in (28). Table D2 in Appendix D.2 reports gravity estimates for trade flows relevant to our sample, made of 39 African countries observed over the period 1970-2018. We interact each of the trade impediments (distance, political border and RTA) with a dummy coding for intra-African trade. We obtain $\mu_1 = -0.85$, $\mu_2 = -5.34$ and $\mu_3 = 1.71$. Finally, the trade elasticity is calibrated based on the meta-analysis in Head and Mayer (2014) and set to $1 - \sigma = -5.03$. The overall calibration of τ_{in} is achieved by incorporating all these elements into equation (28).

Step 6: The final step is to recover farming productivity $\widehat{A_i}$ for all regions. This is accomplished using equation (13), where unobservables are replaced with parameters retrieved from the first five steps, allowing us to manipulate and express it as:

$$A_i^{1-\sigma} = \sum_{n=1}^{N} \frac{\hat{\tau}_{in}^{-(\sigma-1)} \left(\texttt{light}_i\right)^{\lambda\times(1-\sigma)}}{\sum_{k=1}^{N} \hat{\tau}_{kn}^{-(\sigma-1)} A_k^{\sigma-1} \left(\texttt{light}_k\right)^{\lambda\times(1-\sigma)}} \widehat{s}_n \left(\frac{\texttt{light}_n}{\texttt{light}_i}\right)^{\lambda} \frac{\texttt{pop}_n}{\widehat{L}_i}. \tag{29}$$

To obtain $\widehat{A_i}$, we use an iterative fixed-point procedure: starting with an initial vector of A_i, which we set as the vector of nighttime lights, one can compute the right-hand side of equation (29), which gives a new value for the each of i's productivity, A_i'. We use a dampening factor δ to update the new vector of productivity as $\delta \times A_i + (1 - \delta) \times A_i'$. Iterating those steps until the vector converges provides a fixed-point vector $\widehat{A_i}$. To cross-validate the method, we plot $\widehat{A_i}$ against observed nighttime light (Figure D2). As expected, higher levels of nighttime light—a proxy for wages in our procedure—are associated with larger estimates of the farming productivity.

5 Counterfactual Simulation

5.1 Methodology

Our approach to counterfactual analysis involves changing several exogenous variables, particularly bilateral trade barriers and/or productivity across ethnic regions in Africa, and then computing the new equilibrium to compare with the initial one. This approach has been labeled the Difference in Expected Values (DEV) approach to counterfactual analysis by Head and Mayer (2019), since it relies on what *should be* the equilibrium in expectation according to the model evaluated at observed values of covariates and estimated parameters.[29] We therefore have to first compute the

[29] This approach is appropriately called the "covariates-based" approach by Dingel and Tintelnot (2020). An alternative is to leverage the multiplicative nature of the model to predict changes in the main outcomes when a policy change is imposed on the *observed values* of those outcomes. This is called the Exact Hat Algebra approach in the literature since Dekle et al. (2007)—the "calibrated-shares procedure" for Dingel and Tintelnot (2020). While the debate regarding the respective merits of the two approaches is still ongoing, we do not have the luxury of a choice in our case. The Exact Hat Algebra approach relies on observing market shares, which we cannot implement in our context as we do not observe shares of trade flows (π_{in}) between the regions of interest in Africa.

expected *factual* equilibrium, which differs from the one actually observed in the data, and compare it to the expected *counterfactual* equilibrium after the policy shock.

Section 4 provides the steps to estimate and "invert" the model, such that we recover all parameters governing the factual equilibrium. This set of structural parameters (omitting hat notation to avoid cluttering notation) is given by:

$$\Theta = \{\gamma, \sigma, s_n, \psi_i, \texttt{pop}_i, A_i, \tau_{in}, \xi_{in}\}.$$

The equilibrium vector of wages conditional on Θ is characterized by equation (15) and obtained numerically by a fixed point iteration of the following equation which transforms an initial vector of wages $\mathbf{w}$ into a final vector $\tilde{\mathbf{w}}$:

$$\tilde{w}_i = \sum_{n=1}^{N} \beta_{in}(\mathbf{w}) w_n \frac{\texttt{pop}_n}{\texttt{pop}_i}, \tag{30}$$

where $\beta_{in}(.)$ is a non-linear function of the wage vector $\mathbf{w}$ given by equation (16). The algorithm used to find a numerical solution is as follows:

Wage (inner) loop steps:

Step 0: Use $w_i = \frac{A_i}{A_1}$ as values for initializing the procedure.

Step 1: With the vector of w_i (normalized such that $w_1 = 1$) obtained from the previous iteration, compute the RHS of equation (30). The LHS of equation (30) yields a new vector $\tilde{w}_i$ (also to be normalized by $\tilde{w}_1$).

Step 2: Use a dampening factor δ to compute the new vector of wages as $\delta w_i + (1-\delta)\tilde{w}_i$.

Steps 1 and 2 are repeated until the vector of wages converges within a specified tolerance level. The fixed-point iteration on equation (30) is sufficient for computing the *factual* equilibrium, i.e. the equilibrium values of endogenous variables at baseline values of Θ. Once equipped with the equilibrium wage vector $\mathbf{w}$, we compute the regional income $Y_n = w_n \texttt{pop}_n$; then, equation (8) gives the optimal number of fighters allocated to each battlefield, l_{in}. The remaining endogenous variables can be derived immediately from the equations in Section 2.3. The approach for the counterfactual equilibrium is similar. After modifying one or several structural parameters in Θ, we perform the algorithm again on equation (30) to generate new counterfactual outcomes which we then compare to the factual ones.

5.2 Destruction and security spillovers

Given the substantial evidence regarding the large impact of conflict on output and income, it seems unrealistic to hold the vectors $\mathbf{A}$ and $\mathbf{s}$ constant in our simulations. Violence experienced in a region is highly likely to adversely affect workers' productivity. Furthermore, the share of income secured from looting is influenced, at least partially, by the efficiency of local police and army

forces, as well as broader property rights enforcement mechanisms, which are often facilitated by local income levels. In the counterfactual simulations presented below, the model is therefore augmented with one "destruction spillover":

$$\log A_n = \log \bar{A}_n - \varepsilon_1 \times \texttt{violence}_n \left[\mathbf{w}(\mathbf{A}, \mathbf{s})\right], \tag{31}$$

and one "security spillover":

$$\log \frac{s_n}{1 - s_n} = \log \frac{\bar{s}_n}{1 - \bar{s}_n} + \varepsilon_2 \times \log Y_n \left[\mathbf{w}(\mathbf{A}, \mathbf{s})\right], \tag{32}$$

where parameter $-\varepsilon_1 < 0$ denotes the semi-elasticity of TFP with respect to violence, and $\varepsilon_2 > 0$ is the elasticity of state capacity to income (the log odds ratio form ensuring that it stays within the 0-1 range).[30]

Note that, as soon as productivity depends on violence, solving for the vector of wages remains necessary, but no longer sufficient, to compute equilibrium. This is because wages are both affected by productivity (equation 16), and feedback through equilibrium violence (equation 31). The same applies to state capacities (equations 16 and 32). Therefore, in our counterfactual simulations, we employ an inner-middle-outer loop fixed-point procedure. The inner loop characterizes the equilibrium wage vector $\mathbf{w}$, given productivity $\mathbf{A}$ and state capacity $\mathbf{s}$ vectors (as described above). The middle loop takes the resulting $\mathbf{w}$ vector from the inner loop to determine the equilibrium state capacity vector $\mathbf{s}$, while the outer loop takes the equilibrium $\mathbf{s}$ and finds $\mathbf{A}$:

State capacity (middle) loop steps:

Step 1: Use the value of w_n obtained from the inner loop and insert $Y_n = w_n \texttt{pop}_n$ in the RHS of equation (32). The LHS of equation (32) yields a new log odds ratio that enables to recover a new vector $\tilde{s}_n$.

Step 2: Use a dampening factor δ to compute the new vector of state capacities as $\delta s_n + (1 - \delta)\tilde{s}_n$.

Steps 1 and 2 are repeated until the state capacity vector converges within a specified tolerance level.

Productivity (outer) loop steps:

Step 1: Use the vector of s_n (obtained from the middle loop) and w_n (from the inner loop) in equation (18) in order to compute the RHS of equation (31). The LHS of equation (31) yields a new vector $\tilde{A}_n$.

[30] Calibration of ε_1 is based on a 2SLS regression of the log of A_n on $\texttt{violence}_n$, with violence being instrumented by a model-driven supply shifter of violence. The calibrated semi-elasticity of TFP to violence is found to be $-\varepsilon_1 = -0.00044$; a straightforward quantitative interpretation is that one hundred additional ACLED events reduce local TFP by 4.4%. Regarding the dependence of s_n on income, the instrument used to estimate the 2SLS regression based on equation (32) is a model-consistent shifter of trade revenues. We obtain $\varepsilon_2 = 0.5$. The baseline productivity and state capacity parameters, $\bar{A}_n$ and $\bar{s}_n$, are obtained by simple manipulation, once ε_1 and ε_2 are calibrated. See Appendix Section E for more details.

Step 2: Use a dampening factor δ to compute the new vector of productivities as $\delta A_n + (1-\delta)\tilde{A}_n$.

The overall simulation procedure stops when the tolerance level for changes in the vector of **A** is reached.

5.3 Welfare

The welfare impact of the counterfactual scenarios that we simulate can be related to a sufficient statistics formula well-known in trade since the work of Arkolakis et al. (2012). We begin by expressing the equilibrium level of welfare before the policy change, ω_n, as the secured share of income over the CES price index P_n:

$$\omega_n = \frac{s_n w_n}{P_n} \quad \text{and} \quad P_n \equiv \left(\sum_{k=1}^{N} (\tau_{kn} w_k / A_k)^{1-\sigma} \right)^{1/(1-\sigma)}. \tag{33}$$

The share of trade within region n is derived from equation (1) and equal to $\pi_{nn} = (\tau_{nn} w_n / A_n)^{1-\sigma} P_n^{\sigma-1}$ implying that $\frac{w_n}{P_n} = \pi_{nn}^{\frac{1}{1-\sigma}} \frac{A_n}{\tau_{nn}}$. Assuming that τ_{nn} remains unaffected by the counterfactual experiment, the change in welfare between the factual equilibrium (no superscript) and the counterfactual equilibrium (superscript $'$) equals:

$$\frac{\omega_n'}{\omega_n} = \left(\frac{\pi_{nn}'}{\pi_{nn}} \right)^{\frac{1}{1-\sigma}} \times \frac{\bar{A}_n'}{\bar{A}_n} \exp\{-\varepsilon_1 (\texttt{violence}_n' - \texttt{violence}_n)\} \times \frac{s_n'}{s_n}. \tag{34}$$

The first term represents the traditional formula of Arkolakis et al. (2012), which embeds an indirect influence of violence on welfare through changes in trade patterns. The second and third terms are additional direct effects factoring in the destruction and security spillovers, respectively. They reflect equilibrium changes in violence and income, which in turn lead to alterations in productivity and state capacity.

5.4 Factual equilibrium

Table 3 reports a number of important statistics predicted by the model in the initial (factual) equilibrium. For each panel, the three rows report three spatial categories: i) flows that occur within the same ethnic region; ii) within the same country (but between different ethnic regions); and iii) across different countries. The two first columns show the model's predictions for trade and violence, respectively, while the last column reports the observed figures for violence. Finally, panel (a) reports the shares (in percentage) of total trade and violence that take place within each of the three categories. Panel (b) reports the shares of those flows across destinations.

Around 31% of trade occurs within an ethnic region ($i = n$) in our model, and less than 7% occurs between regions belonging to different countries, leaving 62% of trade to take place between different regions of a country. Similar patterns emerge for violence. First, the majority of violence is not perpetrated locally. Internal ("self-inflicted") violence represents less than 22% of the total

Table 3: Factual equilibrium: some statistics

Panel (a)	Total trade	Total violence	
	Pred.	Pred.	Obs
same ethnic region	31.18	21.16	26.79
same country	62.28	50.11	46.37
different countries	6.54	28.72	26.84
Panel (b)	π_{in}	$\frac{\texttt{violence}_{in}}{\texttt{violence}_{n}}$	
	Pred.	Pred.	Obs
same ethnic region	15.10	6.53	6.52
same country	76.54	57.25	57.31
different countries	8.64	36.50	36.45

Note: All figures represent percentage points. Columns 1 and 2 report the levels of trade and violence that the model predicts to take place in the factual equilibrium in each of the categories. Column 3 reports observed violence. Panel (a) reports the shares of total trade and violence that take place within each of the three categories. Panel (b) reports the shares across destinations n.

violence predicted by the model. Secondly, most of the projected spatial violence occurs within country boundaries. Both patterns align well with the gravity results in Section 3.4.

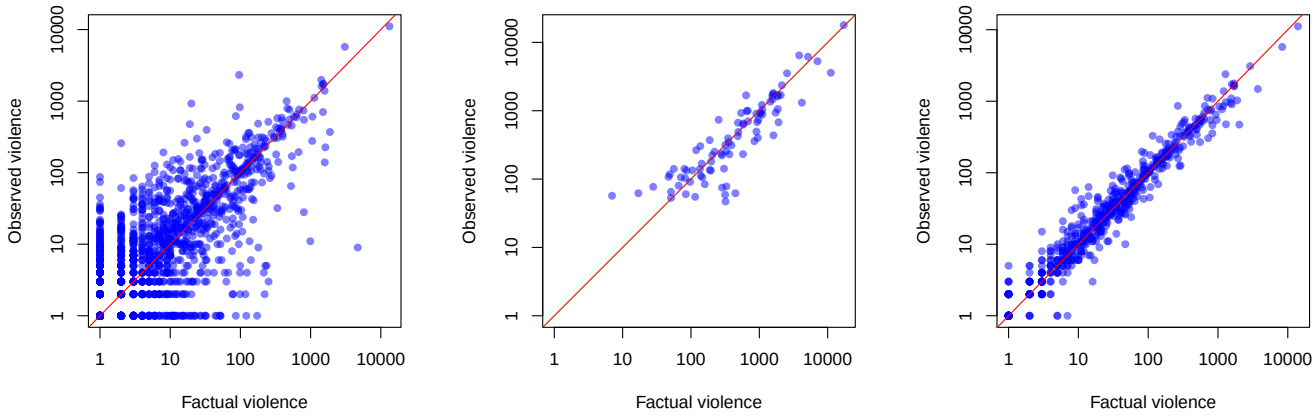
Note that the trade patterns are entirely driven by the structure and parameters imposed on the model (step 5 in Section 4), since trade flows are not observed at that level of detail (i.e. across ethnic groups). However, for bilateral violence, we can confront the levels predicted by the model in column 2 with the observed flows in column 3. The prediction seems to fit the data patterns reasonably well. Figure 4 pushes further the comparison for the whole cross-section. Panel (a) reports on the x-axis bilateral flows of violence between i and n, while the y-axis reports the observed corresponding flow (in log scales and rounded to the nearest unit in both cases). Although deviations are larger for smaller cells where idiosyncratic shocks to violence matter more, as expected, the model performs well in predicting the spatial pattern of violence. Total violence perpetrated by i (panel b) or suffered by n (panel c) is even better predicted.

5.5 Counterfactual 1: A rise in the Kanuri region's productivity

Our first counterfactual explores the consequences of a 10% increase in the baseline productivity ($\bar{A}_n$ in equation 31) of a single region. This exercise serves two objectives: it provides a numerical illustration of the theoretical mechanisms operating in our model and their interactions, and it also addresses one of the most debated questions in the conflict literature regarding how income shifters influence violence. We choose the Kanuri region where our procedure allocates Boko Haram, a notoriously violent group active in Western Africa (see footnote 34). Table 4 summarizes results for three sets of regions: the one where the productivity shock is applied ("shock"); the subset of regions within a 1000km radius ("close to shock"); and the rest of Africa ("other").[31] The columns

[31] All counterfactuals are run for a restricted set of regions within 2000 km of the shock to ease computation and reduce convergence time. The large spatial decay estimated for violence and trade implies that the impact of a shock is almost

Figure 4: Goodness of fit



(a) Bilateral (b) Outward (c) Inward

Note: The x-axis displays the predicted flows of violence (log scale). The y-axis reports the observed flows of violence (log scale). In panel (a) the bilateral flows of violence between i and n are reported. In panel (b) the total flow of violence perpetrated by i is reported. In panel (c) the total flow of violence suffered by n is reported.

provide six outcomes averaged over the regions in those three sets: change in local violence, change in state capacity, change in the number of fighters, and the changes in traditional economic variables: TFP, income and welfare. In summary, increases in the first three indicate negative outcomes, while increases in the last three indicate positive outcomes.

We run five types of counterfactuals, gradually incorporating the rich interactions that our model captures between the trade/production part of the economy and the violent part of this world:

1: Pure-trade In this simple scenario, the allocation of labor between production and fighting is held constant.[32] Violence is therefore unaffected by the shock, and only trade patterns exhibit reactions. Because of the endogenous redirection of trade flows, an exogenous 10% increase in TFP translates into a less than proportional increase in income and welfare.[33]

2: Spillover-free. This scenario allows the endogenous reallocation of the labor force between farmers and fighters, which renders violence endogenous in the first column, while keeping the two spillover channels (on A_n and s_n) muted. The rise of TFP in the Kanuri region has two main effects on violence received by the Kanuri: there is a fall in internal violence because w_i rises in equation (8) (Kanuri groups are more attracted to farming, and exert less violence, including locally), compensated by an increase in Y_n (caused by w_n), which attracts violence from everywhere (including locally). The second effect dominates resulting in an increase in received violence. However, the reallocation of the population to farming in the Kanuri region (with the number of fighters

completely muted at greater distances.

[32] Details available in Appendix Section F.

[33] This can be seen by rearranging the goods market clearing equation (13) to isolate nominal wages in i, such that $w_i = A_i^{\frac{(\sigma-1)}{\sigma}} L_i^{-1/\sigma} \sum_{n=1}^{N} \tau_{in}^{-(\sigma-1)} s_n w_n \bar{L}_n P_n^{\sigma-1}$. The proportionality factor to a first order is therefore $\frac{(\sigma-1)}{\sigma} < 1$.

Table 4: Aggregate results across CF scenarios

		Avg. % Change in					
Scenario	Zone	$\texttt{violence}_n$	$1-s_n$	l_n	TFP (A_n)	Income	Welfare
1:Pure Trade	shock	0	0	0	10	7.99	8.14
	close to shock	0	0	0	0	0.01	0.02
	other	0	0	0	0	-0.00	0.00
2:Spillover-free	shock	0.64	0	-7.43	10	8.28	8.46
	close to shock	-2.9	0	0.8	0	-0.01	0.02
	other	-0.39	0	0.15	0	-0.02	0.00
3:Security-inclusive	shock	-3.2	-3.81	-8.5	10	8.26	8.55
	close to shock	-2.89	0.01	0.72	0	-0.01	0.02
	other	-0.38	0.01	0.15	0	-0.02	0.00
4:Damage-inclusive	shock	0.59	0	-6.7	9.51	7.89	8.09
	close to shock	-2.77	0	0.74	0.06	0.04	0.09
	other	-0.39	0	0.22	0.01	-0.02	0.01
5:All-inclusive	shock	-4.26	-5.08	-10.88	13.59	11.22	11.65
	close to shock	-3.84	-0.04	0.8	0.09	0.06	0.13
	other	-0.52	0.01	0.27	0.01	-0.02	0.02

Note: Numbers represent average percent changes. "Shock" = Kanuri Region; "close to shock" = Regions within 1000km from Kanuri; "other" = Regions within 1000km-2000km from Kanuri. The shock is a 10% in increase in the baseline TFP ($\bar{A}_n$ in equation 31) in the Kanuri region. $\texttt{violence}_n$ is the total violence in n. $1-s_n$ stands for the state capacity in n. l_n the number of fighters in n. TFP (A_n)is the productivity in n. Total "Income" of a region is defined in equation (11). "Welfare" refers to the secured per-capita real wage as defined in equation (33).

dropping by around 7%) boosts output and reduces the locally perceived price index. This is reflected in the (modest) increase of 0.3 percentage points of income and welfare gains compared to the "Pure Trade" scenario. The neighboring regions experience a decrease in their incoming violence (mostly from the Kanuri region), but a slight increase in their fighters' share of population (mostly attracted by increased Kanuri income). Overall, these effects offset each other in terms of income and welfare.

3: Security-inclusive. In this scenario, we unleash the security spillover, in which increased income shrinks the unsecured share of that income ($1 - s_n$ decreases by almost 4%). The main consequence of adding this channel is a 3% decrease in incoming violence in the Kanuri region compared to the status quo. This contributes to a slight increase in local welfare with respect to the previous scenario.

4: Damage-inclusive. In this scenario, we replace the security spillover with the damage spillover channel, where violence retroactively affects productive activity. Compared to the "Spillover-free" scenario, the increased violence received by the Kanuri region now reduces TFP by around 0.5 percentage points. As a consequence, local income and welfare do increase, but by a smaller amount than when spillovers are deactivated. Interestingly, the TFP of nearby regions also rises as they experience a reduction in violence. Therefore, their consumers also benefit from the productivity increase in the Kanuri region. This spatial diffusion of the TFP shock happens entirely through the reduction in violence, which is a novel feature of our model.

5: All-inclusive. In the final scenario, both security and damage spillovers are allowed to operate. Their combined action magnifies the outcomes, resulting in a substantial decrease in incoming violence in the Kanuri region, further amplifying the initial productivity impulse, and ultimately increasing the final welfare benefits by almost 40% with respect to the "Spillover-free" scenario ($11.65/8.46 = 1.38$). Thus, the security channel reduces the incentives for violence, leading to less destruction and a subsequent rounds of productivity increases, which in turn boost income and further reinforces security, and so on. These feedback loops between the two spillovers are a novel aspect of our theoretical framework, modeling some elements of the logic underlying the so-called conflict trap in the literature (for a survey, see Rohner and Thoenig, 2021).

Welfare changes: Figure G4 in Appendix provides a decomposition of welfare effects in this "All-inclusive" scenario, where all channels are active. The decomposition follows equation (34), and plots the three elements of the overall welfare change in percentage points against the distance from the shock in kilometers (up to 1000 kms, since it is very clear from the last row of the last column of Table 4 that welfare changes are quite negligible beyond that range). Total welfare change tends to fall sharply with distance from the original shock, as expected, with some nearby regions still experiencing substantial overall gains around 1 to 2%. As is most visible for the Kanuri region, welfare change mostly comes from the increase in TFP. Indeed, the self-trade channel even acts as

a negative contribution, reflecting the shift towards local varieties in the new equilibrium. The rise in s_n has a *direct* but marginal contribution to welfare, due to all regions having initial s_n values close to 1 (including the Kanuri region). However, its indirect effect through the feedback loops highlighted above is critical for welfare, as these interactions between the two spillovers substantially increase the overall TFP compared to the "Damage-inclusive" scenario, where s_n is constant. Figure G5 compares the overall welfare gains in our model to the famous formula from Arkolakis et al. (2012). In panel (a), scenarios 1 and 2 are reported. Since these scenarios lack spillover channels, welfare changes align exactly with the ACR formula, except for the Kanuri region, which shows significant gains coming from the direct TFP shock. Panel (b) illustrates scenario 5, where our spillover channels introduce notable deviations from the ACR formula, even for regions without direct TFP gains. This deviation is attributed to the indirect benefits of reduced violence. The statistic of change in self-trade relevant in the traditional trade model is not sufficient anymore in our model which accounts for endogenous violence.

Geography of violence: Figure 5 provides further details regarding the impact of the productivity shock on the geography of violence between the Kanuri and neighboring regions (all figures are for the "All-inclusive" scenario). The gravity of violence equation (8) yields the following decomposition:

$$\Delta \log \texttt{violence}_{in} = -\frac{\gamma}{1-\gamma}\Delta \log \xi_{in} - \frac{1}{1-\gamma}\Delta \log w_i + \Delta \log w_n + \Delta \log(1-s_n) - \Delta \log \text{MRV}_n, \quad (35)$$

with

$$\text{MRV}_n = \xi_{in}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_i}{w_i}\right)^{\frac{\gamma}{1-\gamma}} + \sum_{k\neq i} \xi_{kn}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_k}{w_k}\right)^{\frac{\gamma}{1-\gamma}},$$

which can be used to study the impact of any shock on each bilateral flow of violence. Here, we analyze the consequences of the 10% productivity increase in the Kanuri region on (i) total violence taking place in the Kanuri region (due to changes in internal as well as in incoming violence), and (ii) violence flowing to other regions. The analysis relies on the theoretical channels discussed in Section 2.2.

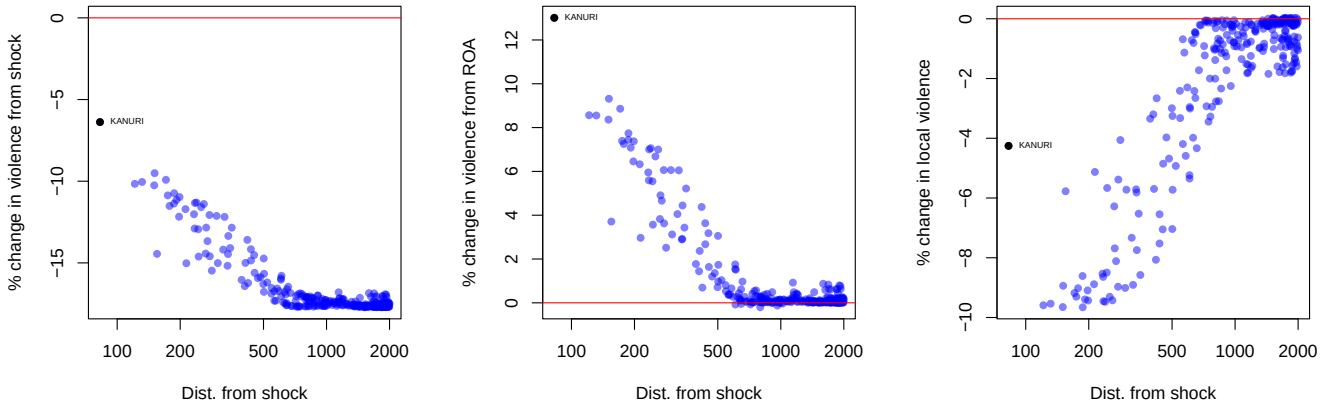
As noted above, the increase in Kanuri's productivity leads to a rise in their local wage (see footnote 33 for details). A higher wage decreases violence perpetrated by $i =$ Kanuri (*opportunity cost* channel) and increases violence targeting $n =$ Kanuri (*rapacity* channel). Additionally, there is a decline in $(1 - s_n)$ due to security spillovers, which reduces violence targeting Kanuri (*state capacity* channel).

(i) What is the consequence for Kanuri's internal violence? For $i = n$, the changes in wage act in opposite direction such that the *rapacity* and *opportunity cost* channels tend to offset each other. Inspection of equation (35) reveals that cancellation is exact in the limit when violence is very local (see also footnote 13) and only the *state capacity* channel matters. This logic is at work in panel (a) of Figure 5, where internal Kanuri violence falls by a little more than 5%. In

panel (b), we see the other side of the coin: the increased income of the Kanuri incentivizes rapacity, leading neighboring groups to deploy more fighters to the Kanuri region to seize that additional income. This mechanism is reinforced by the fact that the Kanuri perpetrate less violence, thereby reducing MRV_n and competition in the appropriation market. Overall, the decrease in internal violence outweighs the increase in inward violence, resulting in a net reduction in the total level of violence in the Kanuri region by slightly more than 4% (the black circle in panel (c) of Figure 5, also visible in first column and first row of "All-inclusive" scenario in Table 4).

(ii) What happens to violence perpetrated in other regions? Equation (35) highlights how competition in the appropriation market, as captured by MRV_n, shapes the spatial reshuffling of violence. Returning to Figure 5(a), we observe that, initially, the rise in Kanuri's wage reduces its violence towards every other region. However, this effect is not uniformly distributed: a spatial gradient is visible. In regions very far from the Kanuri region, only the $-\frac{1}{1-\gamma}\Delta \log w_i$ term in equation (35) is active since $\xi_{in}^{-\frac{\gamma}{1-\gamma}}$ approaches 0 in MRV_n. Conversely, as a region n gets closer to the Kanuri region, the countervailing term involving w_i in MRV_n becomes stronger, dampening the violence-reducing effect of Kanuri's TFP increase. A related spatial gradient is visible in panel (b): because the Kanuri group is less active in regions close to it, those regions receive more attacks from other region in Africa. This substitution effect is driven by the change in MRV_n. In panel (c) the net effect is revealed to be violence-reducing for those regions surrounding the positive TFP shock.

Figure 5: Change in violence after 10% rise of productivity in Kanuri's region



(a) Violence from Kanuri $\texttt{violence}_{in}$, i=Kanuri

(b) Violence from RoAfrica $\sum_{k \neq i} \texttt{violence}_{kn}$

(c) Violence received $\texttt{violence}_n$

Note: The x-axis displays the distance from the 10% increase in productivity in the Kanuri region. The y-axis reports the overall violence change in percentage points. All figures come from the "All-inclusive" scenario. In panel (a), the flows of violence from Kanuri are reported. In panel (b), the flows of violence from all African regions are reported. In panel (c), the flows of violence received by region n are reported.

5.6 Counterfactual 2: Sahelxit

Our second counterfactual analysis quantifies the changes in violence and development in West Africa expected from the "Sahelxit." This event occurred on January 28th, 2024, when Burkina Faso, Mali, and Niger withdrew from the CEDEAO trade agreement with immediate effect. These three exiting countries share several characteristics: they are located in the Sahel, are former French colonies, are landlocked, are among the world's poorest nations, face significant pressure from transborder armed groups, and have been under military rule since 2022, 2021, and 2023, respectively. While the adverse economic effects expected from such trade disintegration are relatively well understood, less is known about its geopolitical impact—a crucial issue given the already high prevalence of violence in the Sahelian region. Our model sheds light on this question by simulating the complex general equilibrium interactions between trade-induced income changes and violence.

Context: In Figure 6, panel (a), the gray areas depict the 243 ethnic regions primarily affiliated with a pre-exit CEDEAO member country. Among those, Sahelxit regions, namely the ones located within Burkina, Mali or Niger, are shaded in dark gray. CEDEAO, overall, exhibits a high propensity for violence, hosting 26 origins of violence (panel b) and 195 destinations of violence (panel c).[34] The "Sahelexit" regions, although comprising only 14.85% of the population of the CEDEAO region, accounts for a disproportionate 36.49% of the total violence committed by actors located in the CEDEAO. Additional contextual elements are reported in Appendix Section G.2

Implementation and results: Our Sahelxit scenario amounts to terminating the existing Regional Trade Agreement (RTA) between Burkina-Faso, Mali and Niger and the rest of CEDEAO. We use our model to compute the new general equilibrium after imposing an increase in trade frictions between all pairs with i belonging to one of the three exiting countries and n in the rest of CEDEAO. More specifically, for these "treated" pairs *only*, we remove the RTA effect in equation (28). This is equivalent to imposing the following shock to their trade frictions:

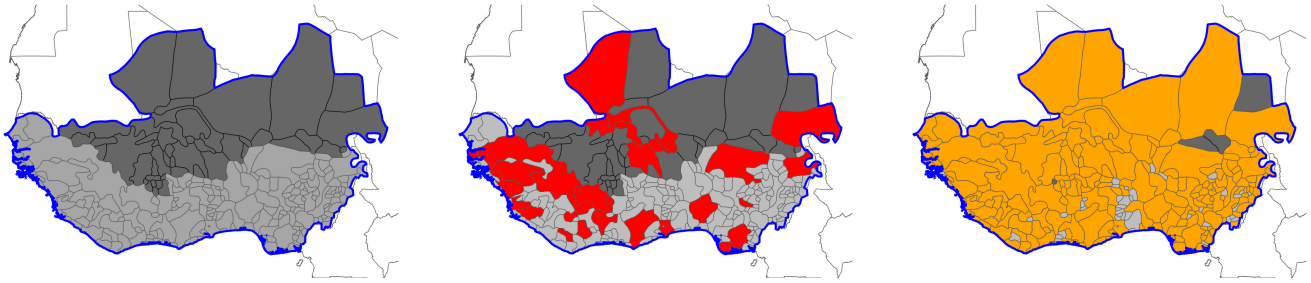
$$\tau'_{in} = \tau_{in} \times e^{\mu_3/(\sigma-1)} = \tau_{in} \times e^{1.71/5.03},$$

where τ'_{in} stands for their post-sahel-exit friction. Note that this is a significant shock, as trade flows between the Sahel and rest of CEDEAO are predicted to fall by approximately $82\% = \exp(-1.71) - 1$.

Table 5 provides a number of summary results regarding bilateral trade and bilateral violence changes due to the simulated policy experiment. The three rows detail results according to three categories of destinations: those breaking away from CEDEAO; those staying in CEDEAO; and the other regions in Africa. The four columns on the left of this table examine changes in violence over origins, and the four to the right report changes in bilateral trade shares. Table 6 reports the

[34] In particular, two of Africa's most violent actors are active in these territories: Jama'at Nusrat al-Islam wal-Muslimin (JNIM), the second most violent actor (4223 events), and Boko Haram, the eighth most violent actor (1605 events). Both are associated with ethnic groups located in the CEDEAO region: the Masina ethnic group for the former and the Kanuri ethnic group for the latter.

Figure 6: Geography of violence in Sahel and neighbouring countries



(a) Sahel exiters and CEDEAO (b) 26 Origins of violence (c) 195 Destinations of violence

Note: In panel (a), the light gray areas depict the 243 ethnic regions primarily affiliated with a pre-exit CEDEAO member country. Sahelxit regions are shaded in dark gray. Panel (b) depicts the origins of violence, and panel (c) shows the destinations of violence.

Table 5: Aggregate results of Sahelxit scenarios: dyadic effects

	Zones	Avg. % change in $\texttt{violence}_{in}$				Avg. % change in π_{in}			
	Origin →	Local	Rest of :			Local	Rest of :		
Scenario	Destination ↓		sahelxit	CEDEAO	Africa		sahelxit	CEDEAO	Africa
All-inclusive	sahelxit	0.61	0.8	-0.89	-1.04	9.51	8.43	-81.18	3.69
	rest of CEDEAO	0.04	1.76	0.06	-0.1	1.40	-80.91	1.26	0.89
	other	-0.03	1.75	0.04	-0.11	-0.02	4.69	0.46	0.09

Note: Numbers represent average percent changes. The average is taken over the ethnic regions inside a destination zone.

Table 6: Aggregate results of Sahelxit scenario: monadic effects

		Avg. % Change in					
Scenario	Zone	$\texttt{violence}_n$	$1 - s_n$	l_n	TFP (A_n)	Income	Welfare
All-inclusive	sahelxit	0.24	0.47	0.77	-0.03	-0.96	-1.69
	rest of CEDEAO	0.12	0.03	-0.01	-0.00	-0.07	-0.24
	other	0.17	-0.01	-0.36	-0.00	0.02	0.01

Note: Numbers represent average percent changes.

changes in the main monadic variables (total violence, insecurity, number of fighters, TFP, income and welfare) for the same three categories of regions. In both tables, the results are computed for the "full" version of the model, accounting for both security and damage spillovers.

(i) Due to the increase in trade frictions, trade is redirected away from the pairs of regions where the RTA link is shut down. For the most affected regions (Sahelxit), the reallocation of trade is local (within the region itself) or with other Sahelxit regions. The rest of CEDEAO regions also see their local trade increase. In the rest of Africa, most of the action is an increase in imports from Sahelxit regions, which stems from those regions becoming poorer and therefore more competitive in exporting to the rest of Africa. This can be seen in the first row of Table 6, where income in Sahelxit regions drops by around 1%.

(ii) For Sahelxit regions, the end of the RTA with the rest of CEDEAO represents a significant loss of economic opportunities. The implication for violence is that looting becomes relatively more attractive as an occupation, drawing more individuals into fighting. Table 6 shows that the number of fighters goes up by 0.77%. As a result, Sahelxit regions increase their fighting activities across all regions (Table 5): violence increases by 0.61% internally, 0.8% towards other Sahelxit regions, and 1.7% towards the rest of CEDEAO and the rest of Africa. Conversely, as Sahelxit regions become poorer, their attractiveness for rapacity decreases, leading to a reduction in violence targeting them from the rest of CEDEAO (-0.89%) and Africa (-1%).

To further understand the reshuffling of violent flows, it is useful to quantify the different elements of the decomposition outlined in equation (35). This is done in Table 7, which focuses on bilateral flows of violence received by Sahelxit regions from all the CEDEAO. It is decomposed into three categories of origins: rest of CEDEAO (first row); other Sahelxit regions (second row); self (third row). We discuss in details only the first case but the same logic applies to the other two.

(iii) Violence originating from the rest of CEDEAO towards Sahelxit regions decreases on average by 0.89%. This is because the fall in w_n makes looting in those destinations less attractive (*rapacity* effect). The MRV_n term acts similarly, reflecting that the larger number of internal fighters in Sahelxit regions intensifies competition for income appropriation. However, these effects are partially counterbalanced by two forces. First, the rise in $1 - s_n$ (*state capacity* channel), since poorer Sahelxit regions have greater difficulty protecting their income from looting. Second, the decrease in w_i (*opportunity cost* channel) as the rest of CEDEAO also becomes marginally poorer (row 2, column 5 of Table 6). Overall, the pacifying effects largely dominate for this flow of violence. However, as shown in the second and third rows of Table 7, the balance of effects shifts when considering violence originating from Sahelxit regions and self, where the fall in w_i incentivizes fighters and outweighs the previously mentioned pacifying effects.

Figures G6 and G7 in Appendix G.2 illustrate the changes in income and violence across ethnic regions in West Africa. Although significant inter-regional heterogeneity persists even after

accounting for the extent of exposure to the Sahelxit trade shock, visual inspection confirms the patterns observed in the aggregate sample. The Sahelxit leads to a substantial decrease in income within regions that break away from CEDEAO. These regions exert more violence everywhere, while other regions exert less. In net, the level of violence increases in almost all regions in West Africa. Consequently, the Sahelxit is likely to further contribute to the geopolitical destabilization of this part of the continent.

Table 7: All-inclusive CF: decomposition of bilateral violence ($n \in$ Sahelexit)

Origin	Count	Average	Average Δ log					
	dyads	$\texttt{violence}_{in}$	$\texttt{violence}_{in}$	$\xi_{in}^{-\frac{\gamma}{1-\gamma}}$	$w_i^{-\frac{1}{1-\gamma}}$	$(1-s_n)$	w_n	MRV_n^{-1}
Rest of CEDEAO	840	0.69	-0.90	0	0.16	0.47	-0.95	-0.58
Rest of sahelexit	246	24.11	0.80	0	1.86	0.47	-0.95	-0.58
Local	6	354.06	0.60	0	1.86	0.50	-1.02	-0.73
All	1092	7.91	-0.51	0	0.56	0.47	-0.95	-0.58

Note: The sample consists of destinations inside the Sahelexit zone, and origins in the whole CEDEAO. Numbers in last 5 columns represent average percentage point changes. First two columns are mutually exclusive.

6 Conclusion

In this paper, we build a quantitative spatial model of violence, offering a tractable and estimable framework that integrates economic and fighting margins. By leveraging on conceptual and functional form similarities between commonly used trade and conflict models, we demonstrate that gravity forces are useful for the modeling and the description of bilateral violence. Interestingly, our model incorporates the well-known channels through which income influences violence while also uncovering a novel mechanism associated with spatial competition in the appropriation market among fighting groups (the Multilateral Resistance of Violence). Our empirical analysis, covering continental Africa from 1997 to 2023, validates the model's predictions and highlights the significant effects that spatial frictions, such as distance and the crossing of ethnic or political borders, have on conflicts. Overall, our approach is simple, flexible, and requires minimal amounts of data, which we view as a key advantage when analyzing regions in conflict.

Once the structural parameters of our model have been estimated, we can proceed to counterfactual analysis of policy shocks relevant for contexts with weak institutions and where insecurity is prevalent. Our first simulation examines the impact of an exogenous increase in local farming productivity. Accounting for spatial general equilibrium feedback effects significantly increases welfare gains: Indeed, the shock not only reduces violence in the impacted region but also leads to overall pacification and increased productivity in neighboring regions. Our second counterfactual analysis examines the "Sahelxit" event, when Burkina Faso, Mali, and Niger withdrew from the CEDEAO trade agreement on January 28th, 2024. Our simulation predicts that the withdrawal will result in economic decline, increased looting, and more outward violence from these regions, leading to higher overall violence and further destabilization across Western Africa. These counterfactual simulations provide insights to inform policies that pursue the dual objectives of development

and peace. We also perform several decomposition exercises to quantify the contribution of each behavioral margin to the overall effect. Distinguishing between income at origin and destination and factoring in the multilateral resistance of violence (two of our contributions to the literature) are both quantitatively important.

Our paper opens new avenues for future research, and our generic framework can also be further adapted to other contexts. First, the only observable outcome in many settings is the total (monadic) violence exerted in a region, the origin of violence being unknown. Our framework suggests that when estimating the impact of local income on violence in a region without bilateral data, researchers should try to account for the "Supply Potential of Violence". This object encompasses the local opportunity cost of fighting and the weighted average of fighting potentials from all origins, including the destination itself, that seek to appropriate income in the region. Second, our framework can be extended in several ways. For example, climate change is expected to significantly affect local incomes, leading to spatial income inequality and influencing various forms of migration, both internal and international. Our framework should be tractable enough to allow for these effects on income and for associated migration incentives. Another potential extension is to refine the modeling of the industrial organization of armed groups. These groups may adopt diverse organizational structures to sustain their activities and exert control over territory and populations, using both local and global networks for recruitment and operations. Enhancing the representation of the "market power" and various competitive strategies of fighting groups could significantly advance our understanding of how violence develops in a spatially competitive environment.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High Wage Workers and High Wage Firms. *Econometrica 67*(2), 251–333.

Adhvaryu, A., J. Fenske, G. Khanna, and A. Nyshadham (2021). Resources, conflict, and economic development in africa. *Journal of Development Economics 149*.

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

Allen, T., C. Arkolakis, and Y. Takahashi (2020). Universal gravity. *Journal of Political Economy 128*(2), 393 – 433.

Alvarez, F. and R. E. Lucas (2007). General equilibrium analysis of the eaton–kortum model of international trade. *Journal of monetary Economics 54*(6), 1726–1768.

Amarasinghe, A., P. Raschky, Z. Y., and J. Zhou (2020). "conflicts in spatial networks. *mimeo Monash University*.

Anderson, J. E. and E. van Wincoop (2003). Gravity with gravitas: A solution to the border puzzle. *The American Economic Review 93*(1), 170–192.

Anderson, S., A. De Palma, and J. Thisse (1992). *Discrete choice theory of product differentiation*. MIT Press.

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2012). New trade models, same old gains? *American Economic Review 102*(1), 94–130.

Armington, P. S. (1969). A theory of demand for products distinguished by place of production. *Staff Papers - International Monetary Fund 16*(1), 159–178.

Berman, N., M. Couttenier, D. Rohner, and M. Thoenig (2017, June). This mine is mine! how minerals fuel conflicts in africa. *American Economic Review 107*(6), 1564–1610.

Berman, N., M. Couttenier, and R. Soubeyran (2021). Fertile Ground for Conflict. *Journal of the European Economic Association 19*(1), 82–127.

Bruederle, A. and R. Hodler (2018, 09). Nighttime lights as a proxy for human development at the local level. *PLOS ONE 13*(9), 1–22.

Burke, M., S. M. Hsiang, and E. Miguel (2015). Climate and conflict. *Annual Review of Economics 7*(1), 577–617.

Caselli, F. and W. J. Coleman (2013). On the theory of ethnic conflict. *Journal of the European Economic Association 11*(s1), 161–192.

Cederman, L.-E., A. Wimmer, and B. Min (2010). Why do ethnic groups rebel? new data and analysis. *World Politics 62*(1), 87–119.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review 98*(4), 1707–21.

Collier, P. and A. Hoeffler (2004, October). Greed and grievance in civil war. *Oxford Economic Papers 56*(4), 563–595.

Combes, P.-P., M. Lafourcade, and T. Mayer (2005). The trade-creating effects of business and social networks: evidence from france. *Journal of International Economics 66*(1), 1–29.

Crost, B., J. Felter, and P. Johnston (2014, June). Aid under fire: Development projects and civil conflict. *American Economic Review 104*(6), 1833–1856.

Dal Bó, E. and P. Dal Bó (2011, August). Workers, Warriors, And Criminals: Social Conflict In General Equilibrium. *Journal of the European Economic Association 9*(4), 646–677.

Dekle, R., J. Eaton, and S. Kortum (2007). Unbalanced Trade. *American Economic Review 97*(2), 351–355.

Dell, M., B. F. Jones, and B. A. Olken (2014, September). What do we learn from the weather? the new climate-economy literature. *Journal of Economic Literature 52*(3), 740–98.

Dingel, J. and F. Tintelnot (2020). Spatial economics for granular settings. Working Paper 27287, National Bureau of Economic Research.

Dube, O. and J. Vargas (2013). Commodity price shocks and civil conflict: Evidence from colombia. *Review of Economic Studies 80*(4), 1384–1421.

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779.

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

Elbadawi, I. A. and N. Sambanis (2000). External interventions and the duration of civil wars. Technical report, World Bank.

Fally, T. (2015). Structural gravity and fixed effects. *Journal of International Economics 97*(1), 76–85.

Fearon, J. D. and D. D. Laitin (1999). Weak state, rough terrain and large-scale ethnic violence since 1945. *Annual Meetings of the American Political Science Association*.

Fischer, G., F. Nachtergaele, S. Prieler, H. T. van Velthuizen, L. Verelst, and D. Wiberg (2008). Global agro-ecological zones assessment for agriculture (gaez 2008). *IIASA, Laxenburg, Austria and FAO, Rome, Italy.*.

Francois, P., I. Rainer, and F. Trebbi (2015). How is power shared in africa? *Econometrica 83*(2), 465–503.

Gates, S. (2002). Recruitment and allegiance: The microfoundations of rebellion. *The Journal of Conflict Resolution 46*(1), 111–130.

Gates, S. (2017). Membership matters: Coerced recruits and rebel allegiance. *Journal of Peace Research 54*(5), 674–686.

Gehring, K. and P. Schaudt (2024). Insuring peace: Index-based livestock insurance, droughts, and conflict. *CESifo Working Paper No. 10423*.

Guarnieri, E. and A. Tur-Prats (2023). Cultural distance and conflict-related sexual violence. *The Quaterly Journal of Economics 138*(3), 1817–1861.

Harari, M. and E. L. Ferrara (2018, October). Conflict, Climate, and Cells: A Disaggregated Analysis. *The Review of Economics and Statistics 100*(4), 594–608.

Head, K. and T. Mayer (2014). Gravity equations: Workhorse,toolkit, and cookbook. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of international economics*, Volume 4, pp. 131–196. Elsevier.

Head, K. and T. Mayer (2019). Brands in Motion: How frictions shape multinational production. *American Economic Review 109*(9), 3073–3124.

Head, K. and T. Mayer (2021). The United States of Europe: A gravity model evaluation of the four freedoms. *Journal of Economic Perspectives*.

Korovkin, V. and A. Makarin (2021). Production networks and war: Evidence from ukraine. CEPR Discussion Papers 16759, C.E.P.R. Discussion Papers.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *The American Economic Review 70*(5), 950–959.

König, M. D., D. Rohner, M. Thoenig, and F. Zilibotti (2017). Networks in conflict: Theory and evidence from the great war of africa. *Econometrica 85*(4), 1093–1132.

Li, X. and Y. Zhou (2017). A stepwise calibration of global dmsp/ols stable nighttime light data (1992–2013). *Remote Sensing 9*(6).

Lloyd, C. T., H. Chamberlain, D. Kerr, G. Yetman, L. Pistolesi, F. R. Stevens, A. E. Gaughan, J. J. Nieves, G. Hornby, K. MacManus, P. Sinha, M. Bondarenko, A. Sorichetta, and A. J. Tatem (2019). Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. *Big Earth Data 3*(2), 108–139.

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

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

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 (2016). The long-run effects of the scramble for africa. *American Economic Review 106*(7), 1802–48.

Moscona, J., N. Nunn, and J. A. Robinson (2020). Segmentary lineage organization and conflict in sub-saharan africa. *Econometrica 88*(5), 1999–2036.

Mueller, H., D. Rohner, and D. Schönholzer (2022). Ethnic Violence Across Space ['A theoretical foundation for the gravity equation']. *The Economic Journal 132*(642), 709–740.

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

Nunn, N. and N. Qian (2014, June). Us food aid and civil conflict. *American Economic Review 104*(6), 1630–66.

OECD (2009). *Armed violence reduction: Enabling development*. OECD.

Olson, M. (1993). Dictatorship, democracy, and development. *American Political Science Review 87*(3), 567–576.

Proost, S. and J.-F. Thisse (2019, September). What can be learned from spatial economics? *Journal of Economic Literature 57*(3), 575–643.

Redding, S. and E. Rossi-Hansberg (2017). Quantitative spatial economics. *Annual Review of Economics 9*(1), 21–58.

Redding, S. J. (2022). Trade and geography. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of International Economics: International Trade, Volume 5*, Volume 5 of *Handbook of International Economics*, pp. 147–217. Elsevier.

Rohner, D. and M. Thoenig (2021). The elusive peace dividend of development policy: From war traps to macro complementarities. *Annual Review of Economics 13*(1), 111–131.

Sambanis, N. (2001). Do ethnic and nonethnic civil wars have the same causes?: A theoretical and empirical inquiry (part 1). *The Journal of Conflict Resolution 45*(3), 259–282.

Sampson, T. (2017). Brexit: the economics of international disintegration. *Journal of Economic perspectives 31*(4), 163–184.

Sanchez de la Sierra, R. (2020). On the Origins of the State: Stationary Bandits and Taxation in Eastern Congo. *Journal of Political Economy 128*(1), 32–74.

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

Sestito, M. (2024). Identity conflict, ethnocentrism and social cohesions. *Working Paper*.

Tilly, C. (1985). *War Making and State Making as Organized Crime*, pp. 169–191. Cambridge University Press.

Weinstein, J. M. (2005). Resources and the information problem in rebel recruitment. *The Journal of Conflict Resolution 49*(4), 598–624.

Weinstein, J. M. (2007). *Inside Rebellion: The Politics of Insurgent Violence*. Cambridge University Press.

World Bank Group (2020). Commodity markets outlook, october 2020. https://openknowledge.worldbank.org/handle/10986/34621 , License: CC BY 3.0 IGO.

Appendix

A The appropriation game

Setup. Appropriation is modeled as a multi-stage game with repeated contests. The sequence works as follows: farmers are first looted by fighters. Then, fighters are themselves looted by other fighters who are looted by other fighters... until the game ends. The game ends when income is ε-secured (i.e. unsecured income falls below an arbitrary $\varepsilon > 0$) for *all* players.[35] The box below describes the game:

Sequence of the appropriation game:

1. Front-loaded payment of the gross income Y_n^P of producers; only a share s_n is immediately (and definitively) secured. The residual $(1 - s_n)Y_n^P$ is unsecured.
2. "Once-for-all" optimal assignment of fighters l_{in} from region i to region $n \rightarrow$ this sets the stationary appropriation shares p_{in} as defined by the Contest Success Function and subject to a spatial friction factor ξ_{in}.
3. Sub-period 1: Fighters in i loot unsecured farmers' income, generating revenue $R_i(1) = \sum_n p_{in} \times (1 - s_n)Y_n^P$.
4. Repeated stage game starts:

   Stage Game at sub-period $k > 1$:

   i/ Looting by fighters l_{in} of income that is still unsecured in region n.

   ii/ $R_i(k)$ is the flow of income appropriated by fighters in sub-period k. It is (friction-free) repatriated in region i.

   iii/ A share s_i of $R_i(k)$ is *definitively* secured. The residual income $(1 - s_i)R_i(k)$ is unsecured.

   iv/ If $(1 - s_i)R_i(k) < \varepsilon$ for all i, the sub-game ends and we move to stage 5 (below). Otherwise, we proceed to sub-period $(k + 1)$ and restart the stage game (i) to (iii).
5. Secured incomes of all resident workers from i (farmers and fighters) being repatriated in i, production, trade, and consumption take place.

Accounting exercise: "Follow the money". To understand how looting of resources evolves over time, we need to perform an accounting of the looted resources at each stage of the game. The law of motion of appropriation is given by

$$R_i(1) = \sum_n p_{in} \times (1 - s_n)Y_n^P, \quad \text{and} \quad R_i(k) = \sum_n p_{in} \times (1 - s_n)R_n(k - 1), \quad \text{for} \quad k > 1. \tag{A.1}$$

[35] An upper bound of the end game sub-period T is given by $\max_n(1 - s_n)^T < \varepsilon$. In the model inversion procedure, the lowest value of s_n recovered from the data is equal to 0.62. This implies that after 10 iterations of the stage game, each player has at least 99% ($= 1 - 0.62^{10}$) of her income being definitely secured.

In matrix notation:

$$\mathbf{R}(1) = \mathbf{A}\mathbf{Y}^{\mathbf{P}}, \quad \text{and} \quad \mathbf{R}(k) = \mathbf{A}\mathbf{R}(k-1) \text{ for } \quad k > 1, \tag{A.2}$$

where $\mathbf{A}$ is the $(N \times N)$ appropriation matrix: $a_{in} = p_{in} \times (1 - s_n)$.

Replacing $\mathbf{R}(k-1) = \mathbf{A}\mathbf{R}(k-2)$, $\mathbf{R}(k-2) = \mathbf{A}\mathbf{R}(k-3)$, etc. in equation (A.2) yields

$$\mathbf{R}(k) = \mathbf{A}^k \mathbf{Y}^{\mathbf{P}}. \tag{A.3}$$

As k grows large, the vector $\mathbf{R}(k)$ converges to the null vector (this is because all entries of A are positive and below 1).

The amount of *gross fighting revenues* accumulated over the entire game is given by

$$\mathbf{R} = \sum_{k=1}^{\infty} \mathbf{R}(k) = \sum_{k=1}^{\infty} \mathbf{A}^k \mathbf{Y}^{\mathbf{P}} = \mathbf{A}\left(\mathbf{Y}^{\mathbf{P}} + \sum_{k=1}^{\infty} \mathbf{A}^k \mathbf{Y}^{\mathbf{P}}\right) = \mathbf{A}\left(\mathbf{Y}^{\mathbf{P}} + \mathbf{R}\right) = \mathbf{A}\mathbf{Y}, \tag{A.4}$$

where $\mathbf{Y} \equiv \mathbf{Y}^{\mathbf{P}} + \mathbf{R}$ is the vector of total gross incomes (farmers' + fighters' incomes). From the previous equation, we obtain the gross fighting revenues accruing to region i

$$R_i = \sum_n p_{in} \times (1 - s_n) Y_n. \tag{A.5}$$

This revenue is the one that each group seeks to optimize by assigning optimally its fighters l_{in} to each region n (equation 7 in the main text).

From gross income to expenditure. An important distinction in the model is between aggregate gross income Y_n that is made of *both* secured and unsecured income and aggregate expenditure E_n that is made *only* of fully secured income.

Gross aggregate income is given by producers' and farmers' incomes

$$Y_i = Y_i^P + R_i.$$

Total expenditures of producers are given by

$$E_i^P = s_i Y_i^P.$$

For fighters, at each subperiod k of the game, only a share of their newly appropriated flow of revenues is secured and will ultimately contribute to their expenditures. Thus, their expenditure is given by

$$E_i^F = \sum_{k=1}^{\infty} s_i R_i(k) = s_i \sum_{k=1}^{\infty} R_i(k) = s_i R_i.$$

We consequently get that the total expenditure of region i is given by

$$E_i = E_i^P + E_i^F = s_i \left(Y_i^P + R_i\right) = s_i Y_i, \tag{A.6}$$

which corresponds to equation (4) of the main text.

Consequences for the GE. Under these micro-foundations of the appropriation game, we see from the previous relation that the aggregate ratio E_n/Y_n is equal to the exogenous parameter s_n. Essentially, the assumption that "being a fighter does not protect against being looted" implies that this ratio is not affected by the (endogenous) composition of the workforce at equilibrium: it does

not depend on the number of fighters and farmers. Importantly, the relation $E_n = s_n Y_n$ proves to be **very useful** (but not vital) for the GE analysis: it leads to the symmetry in the aggregate trade and fighting revenue equations. This in turn enables us to characterize the GE with the unique fixed point "master" equation (15) in the main text.

General Equilibrium under alternative microfoundations. Let assume alternatively that only farmers are looted: Hence, fighters' income is immediately and fully secured. The rest of the game is unchanged. Clearly, after one step, farmers are looted and the game ends. Moreover, at equilibrium of the labor market, workers are indifferent between farming and fighting. Therefore:

$$s_i \times w_i^P = w_i^F$$

In turn, denoting w_i the fighters' wage at equilibrium, aggregate expenditures are given by

$$E_n = s_n \frac{w_n}{s_n} L_n + w_n l_n = w_n \bar{L}_n$$

And the system of equations that characterizes the GE is given by

$$\frac{w_i}{s_i} L_i = \sum_n \frac{\tau_{in}^{-(\sigma-1)} \left(\frac{s_i A_i}{w_i}\right)^{\sigma-1}}{\sum_k \tau_{kn}^{-(\sigma-1)} \left(\frac{s_k A_k}{w_k}\right)^{\sigma-1}} w_n \bar{L}_n \tag{A.7}$$

$$w_i l_i = \sum_n \frac{\xi_{in}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_i}{w_i}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_k \xi_{kn}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_k}{w_k}\right)^{\frac{\gamma}{1-\gamma}}} \frac{1-s_n}{s_n} w_n L_n \tag{A.8}$$

$$\bar{L}_i = L_i + l_i \tag{A.9}$$

The system of equations (A.7)-(A.8)-(A.9) is not reducible to a unique fixed-point "master" equation anymore. So, the GE model becomes less tractable: in particular, characterizing the existence and uniqueness of the equilibrium is more challenging. This said, numerical simulations are still feasible and the counterfactual analysis could presumably be performed under this alternative model.

B Derivation of the gravity of violence

This section details the computations required to derive the theoretical gravity equation of violence (8). This object represents the partial equilibrium flow of violence (quantity) from i to n. We demonstrate below that it is given by:

$$\texttt{violence}_{in} \equiv \psi_i l_{in} = \xi_{in}^{-\frac{\gamma}{1-\gamma}} \times \left(\frac{\psi_i}{w_i^F}\right)^{\frac{1}{1-\gamma}} \times \frac{(1-s_n) Y_n}{\sum_k \xi_{kn}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_k}{w_k^F}\right)^{\frac{\gamma}{1-\gamma}}}. \tag{B.10}$$

Proof The proof proceeds in three steps:

1. The optimal spatial allocation choice of troops across regions of destination n is given by:

$$R_i \equiv \max_{l_{in}} \sum_n p_{in}\,(1-s_n)\,Y_n, \quad \text{s.t} \quad l_i = \sum_n l_{in}, \tag{B.11}$$

with

$$p_{in} = \frac{(\xi_{in}^{-1}\psi_i l_{in})^\gamma}{\sum_k (\xi_{kn}^{-1}\psi_k l_{kn})^\gamma}, \tag{B.12}$$

under the assumption $0 < \gamma < 1$ (to get a concave maximization problem) which we estimate to be the case in the empirics. The resulting Lagrangian for this optimization problem writes as:

$$\mathcal{L} = \sum_n \frac{(\xi_{in}^{-1}\psi_i l_{in})^\gamma}{\sum_k (\xi_{kn}^{-1}\psi_k l_{kn})^\gamma}(1-s_n)Y_n - \lambda\left(l_i - \sum_n l_{in}\right).$$

Assuming that groups are small enough to neglect their impact on the overall conditions of violence in n, $\frac{\partial\left[\sum_k (\xi_{kn}^{-1}\psi_k l_{kn})^\gamma\right]}{\partial l_{in}} = 0$, we obtain the first-order-conditions:

$$\frac{\partial \mathcal{L}}{\partial l_{in}} = 0 \Longleftrightarrow l_{in} = \left(\lambda \times \frac{\sum_k \xi_{kn}^{-\gamma}\psi_k^\gamma l_{kn}^\gamma}{\gamma \xi_{in}^{-\gamma}\psi_i^\gamma (1-s_n)Y_n}\right)^{\frac{1}{\gamma-1}},$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \Longleftrightarrow l_i = \sum_n l_{in}.$$

Combining the two FOCs and using $l_i = \sum_n l_{in} = \sum_k l_{ik}$:

$$l_i = \sum_k l_{ik} \Longleftrightarrow l_i = \lambda^{\frac{1}{\gamma-1}} \times \sum_k \left(\frac{\sum_j \xi_{jk}^{-\gamma}\psi_j^\gamma l_{jk}^\gamma}{\gamma \xi_{ik}^{-\gamma}\psi_i^\gamma (1-s_k)Y_k}\right)^{\frac{1}{\gamma-1}} \Longleftrightarrow \lambda^{\frac{1}{\gamma-1}} = \frac{l_i}{\sum_k \left(\frac{\sum_j \xi_{jk}^{-\gamma}\psi_j^\gamma l_{jk}^\gamma}{\gamma \xi_{ik}^{-\gamma}\psi_i^\gamma (1-s_k)Y_k}\right)^{\frac{1}{\gamma-1}}}.$$

Plugging back this expression in the first FOC, we obtain:

$$\frac{l_{in}}{l_i} = \left(\frac{\sum_k \xi_{kn}^{-\gamma}\psi_k^\gamma l_{kn}^\gamma}{\gamma \xi_{in}^{-\gamma}\psi_i^\gamma (1-s_n)Y_n}\right)^{\frac{1}{\gamma-1}} \times \left[\sum_k \left(\frac{\sum_j \xi_{jk}^{-\gamma}\psi_j^\gamma l_{jk}^\gamma}{\gamma \xi_{ik}^{-\gamma}\psi_i^\gamma (1-s_k)Y_k}\right)^{\frac{1}{\gamma-1}}\right]^{-1}. \tag{B.13}$$

One can further note that:

$$\left(\frac{\sum_k \xi_{kn}^{-\gamma}\psi_k^\gamma l_{kn}^\gamma}{\gamma \xi_{in}^{-\gamma}\psi_i^\gamma (1-s_n)Y_n}\right)^{\frac{1}{\gamma-1}} = (\gamma \xi_{in}^{-\gamma}\psi_i^\gamma)^{\frac{1}{1-\gamma}} \times \left(\frac{(1-s_n)Y_n}{\sum_k \xi_{kn}^{-\gamma}\psi_k^\gamma l_{kn}^\gamma}\right)^{\frac{1}{1-\gamma}},$$

and :

$$\sum_k \left(\frac{\sum_j \xi_{jk}^{-\gamma}\psi_j^\gamma l_{jk}^\gamma}{\gamma \xi_{ik}^{-\gamma}\psi_i^\gamma (1-s_k)Y_k}\right)^{\frac{1}{\gamma-1}} = \sum_k \left[\left(\gamma \xi_{ik}^{-\gamma}\psi_i^\gamma\right)^{\frac{1}{1-\gamma}} \times \left(\frac{(1-s_k)Y_k}{\sum_j \xi_{jk}^{-\gamma}\psi_j^\gamma l_{jk}^\gamma}\right)^{\frac{1}{1-\gamma}}\right].$$

Thus, defining the useful term Ω as

$$\Omega_n \equiv \frac{(1-s_n)Y_n}{\sum_k (\xi_{kn}^{-1}\psi_k l_{kn})^\gamma}, \tag{B.14}$$

we can rewrite (B.13) as

$$\frac{l_{in}}{l_i} = (\gamma \xi_{in}^{-\gamma} \psi_i^{\gamma})^{\frac{1}{1-\gamma}} \times \Omega_n^{\frac{1}{1-\gamma}} \times \left[\sum_k \left[\left(\gamma \xi_{ik}^{-\gamma} \psi_i^{\gamma} \right)^{\frac{1}{1-\gamma}} \times \Omega_k^{\frac{1}{1-\gamma}} \right] \right]^{-1},$$

Which implies that optimal l_{in} is given by:

$$l_{in} = \frac{\xi_{in}^{\frac{-\gamma}{1-\gamma}} \Omega_n^{\frac{1}{1-\gamma}}}{\sum_k \xi_{ik}^{\frac{-\gamma}{1-\gamma}} \Omega_k^{\frac{1}{1-\gamma}}} \times l_i. \tag{B.15}$$

Plugging back (B.15) into (B.12), one gets the partial equilibrium probability of contest winning, p_{in}^*:

$$p_{in}^* = \psi_i^{\gamma} \xi_{in}^{-\gamma} \times \left[\frac{\xi_{in}^{\frac{-\gamma}{1-\gamma}} \Omega_n^{\frac{1}{1-\gamma}}}{\sum_k \xi_{ik}^{\frac{-\gamma}{1-\gamma}} \Omega_k^{\frac{1}{1-\gamma}}} \times l_i \right]^{\gamma} \times \left[\sum_k \psi_k^{\gamma} l_{kn}^{\gamma} \xi_{kn}^{-\gamma} \right]^{-1}$$

2. The next step is to use the equilibrium p_{in}^* in total looting revenues of i

$$R_i = \sum_n p_{in}^* (1 - s_n) Y_n. \tag{B.16}$$

Noting that ψ_i^{γ}, l_i^{γ} and $\sum_k \xi_{ik}^{\frac{-\gamma}{1-\gamma}} \Omega_k^{\frac{1}{1-\gamma}}$ do not depend on destination n:

$$\begin{aligned}
R_i &= (l_i \psi_i)^{\gamma} \left[\sum_k \xi_{ik}^{\frac{-\gamma}{1-\gamma}} \Omega_k^{\frac{1}{1-\gamma}} \right]^{-\gamma} \times \sum_n \frac{(1 - s_n) Y_n \xi_{in}^{-\gamma}}{\sum_k \psi_k^{\gamma} l_{kn}^{\gamma} \xi_{kn}^{-\gamma}} \times \left(\xi_{in}^{\frac{-\gamma}{1-\gamma}} \times \Omega_n^{\frac{1}{1-\gamma}} \right)^{\gamma} \\
&= (l_i \psi_i)^{\gamma} \left[\sum_k \xi_{ik}^{\frac{-\gamma}{1-\gamma}} \Omega_k^{\frac{1}{1-\gamma}} \right]^{-\gamma} \times \sum_n \Omega_n \xi_{kn}^{-\gamma} \times \left(\xi_{in}^{\frac{-\gamma}{1-\gamma}} \times \Omega_n^{\frac{1}{1-\gamma}} \right)^{\gamma} \\
&= (l_i \psi_i)^{\gamma} \left[\sum_n \Omega_n^{\frac{1}{1-\gamma}} \xi_{in}^{\frac{-\gamma}{1-\gamma}} \right]^{1-\gamma}
\end{aligned}$$

Using the fact that looting revenues are redistributed among fighters, each paid w_i^F, we have $R_i = w_i^F l_i$ and we can solve for l_i:

$$l_i = \frac{\psi_i^{\frac{\gamma}{1-\gamma}}}{(w_i^F)^{\frac{1}{1-\gamma}}} \sum_n \Omega_n^{\frac{1}{1-\gamma}} \xi_{in}^{\frac{-\gamma}{1-\gamma}}. \tag{B.17}$$

This allows to simplify further the expression for the optimal l_{in} in (B.15) to obtain:

$$l_{in} = \frac{\psi_i^{\frac{\gamma}{1-\gamma}}}{(w_i^F)^{\frac{1}{1-\gamma}}} \xi_{in}^{\frac{-\gamma}{1-\gamma}} \Omega_n^{\frac{1}{1-\gamma}} \tag{B.18}$$

3. As can be seen from (B.14), the above equation still has numbers of fighters (l_{in}) on both sides through the Ω terms. The next step is therefore to solve for Ω_n as being functions of w rather

than l. In order to do that, start with definition of Ω_n to obtain

$$\sum_k (\psi_k l_{kn} \xi_{kn}^{-1})^{\gamma} = \frac{(1-s_n)Y_n}{\Omega_n}.$$

Then replacing equilibrium bilateral allocation of fighters from (B.18) into $\sum_k (\psi_k l_{kn} \xi_{kn}^{-1})^{\gamma}$ yields

$$\begin{aligned}\sum_k (\psi_k l_{kn} \xi_{kn}^{-1})^{\gamma} &= \sum_k \left(\psi_k^{\frac{1}{1-\gamma}} (w_k^F)^{\frac{-1}{1-\gamma}} \xi_{kn}^{\frac{-1}{1-\gamma}} \Omega_n^{\frac{1}{1-\gamma}} \right)^{\gamma} \\ &= \Omega_n^{\frac{\gamma}{1-\gamma}} \sum_k \left(\frac{\psi_k}{w_k^F \xi_{kn}} \right)^{\frac{\gamma}{1-\gamma}}. \end{aligned} \tag{B.19}$$

As a result, we obtain a solution for Ω_n which is independent of l_{in}:

$$\Omega_n^{\frac{1}{1-\gamma}} = \frac{(1-s_n)Y_n}{\sum_k \left(\frac{\psi_k}{w_k^F \xi_{kn}} \right)^{\frac{\gamma}{1-\gamma}}}. \tag{B.20}$$

Defining $\texttt{violence}_{\texttt{in}} \equiv \psi_i l_{in}$ and combining (B.18) with (B.20) gives the final equation for gravity of violence:

$$\texttt{violence}_{\texttt{in}} \equiv \left(\frac{\psi_i}{w_i^F} \right)^{\frac{1}{1-\gamma}} \xi_{in}^{\frac{-\gamma}{1-\gamma}} \frac{(1-s_n)Y_n}{\sum_k \xi_{kn}^{\frac{-\gamma}{1-\gamma}} \left(\frac{\psi_k}{w_k^F} \right)^{\frac{\gamma}{1-\gamma}}}, \tag{B.21}$$

which completes the proof.

Finally, we can note that the equilibrium success probability (using B.18 in B.12) is

$$p_{in} = \frac{\left(\frac{\psi_i}{w_i^F \xi_{in}} \right)^{\frac{\gamma}{1-\gamma}}}{\sum_k \left(\frac{\psi_k}{w_k^F \xi_{kn}} \right)^{\frac{\gamma}{1-\gamma}}}. \tag{B.22}$$

Therefore, after optimizing the allocation of fighters over space, we have that

$$\texttt{violence}_{\texttt{in}} = \frac{\psi_i}{w_i^F} p_{in}(1-s_n)Y_n \Rightarrow w_i^F l_{in} = p_{in}(1-s_n)Y_n,$$

which implies that the financial flow linked to bilateral violence simply multiplies lootable income with the probability of success.

C Equilibrium existence and uniqueness

The proof of the existence and uniqueness of the equilibrium vector of wages as a fixed-point of equation (15) follows an approach similar to Allen (2019). It proceeds in two stages. Firstly, we show that our model generates "well-behaved" excess demand functions for goods and violence for each region $i \in N$. Secondly, we invoke standard results from Mas-Colell et al. (1995) on excess demand functions to establish the existence and uniqueness of the equilibrium. A comprehensive proof is provided in the Online Appendix, Section OA4.

From our master equation (15) we define the following excess demand function $Z_i(\mathbf{w})$ for $i \in N$:

$$Z_i(\mathbf{w}) \equiv \frac{1}{w_i}\left(\sum_n \beta_{in}(\mathbf{w}) w_n \bar{L}_n - w_i \bar{L}_i\right), \tag{C.23}$$

where $\beta_{in}(\mathbf{w})$ is given by Equation (16). Reformulating the previous equation leads to a decomposition of the excess-demand function into two sub-components:

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

The sub-components $Z_i^F(.)$ and $Z_i^P(.)$, and by additivity $Z_i(.)$, are all "well-behaved" in the sense that they satisfy the following set of properties:

(P1) $\forall \mathbf{w} \gg 0, \forall i \in N$, $Z_i(.)$ is continuous

(P2) $\forall i \in N$, $Z_i(.)$ is homogeneous of degree zero

(P3) $\forall \mathbf{w} \gg 0$, we have $\sum_{i \in N} w_i Z_i(\mathbf{w}) = 0$

(P4) $\forall \mathbf{w} \gg 0, \exists k > 0$ such that $\forall i \in N, Z_i(\mathbf{w}) > -k$

(P5) Consider any $\mathbf{w} \in \mathbb{R}^{\|N\|}$ such that there exists $(l, l') \in N \times N$ with $w_l = 0$ and $w_{l'} > 0$. Consider any sequence of wages such that $\mathbf{w}^m \to \mathbf{w}$ as $m \to \infty$. Then:

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

(P6) $\forall i \in N, \forall j \neq i, \frac{\partial Z_i(\mathbf{w})}{\partial w_j} > 0$ (Gross Substitute property)

Since the function $Z_i(.)$ satisfies (P1)-(P5), we can apply Proposition 17.C.1 from Mas-Colell et al. (1995) on page 585, thereby proving the existence of an equilibrium vector of wages. Additionally, this function meets the gross substitute property (P6), which is a sufficient condition to apply Proposition 17.F.3 from Mas-Colell et al. (1995) on page 613, thereby proving the uniqueness of the equilibrium vector of wages.

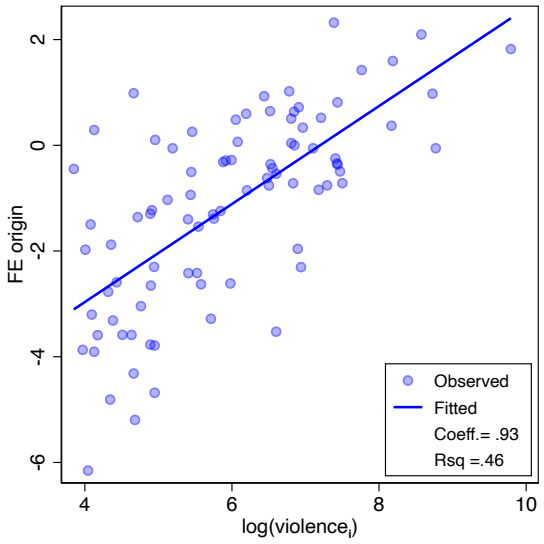
D Supplementary material for section 4 (model inversion)

This section provides supplementary material related to our inversion procedure in Section 4.

D.1 Material for steps 1, 2 and 6

In Figure D1, we plot the estimated $\widehat{\text{FE}}_i^o$ against the observed (logged) number of violent events originating from i. Table D1 reports the estimation results of equation (24) for OLS in column (1), first stage in column (2), reduced form version of the regression in column (3), and 2SLS in column (4). To cross-validate the method to recover farming productivity $\widehat{A_i}$, in Figure D2 we plot (logged) $\widehat{A_i}$ against observed (logged) nighttime light.

Figure D1: Gravity fixed effects and observed violence



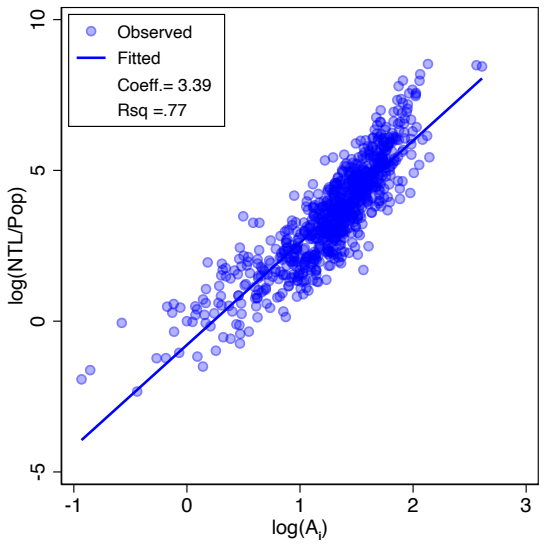
Note: The x-axis displays the observed (logged) number of violent events perpetrated by the 87 origins of violence. The y-axis reports their origin fixed effects, as recovered from the estimation of Equation (22), replicating column 3 of Table 2.

Table D1: Determinants of $\widehat{\text{FE}}_i^o$

	(1)	(2)	(3)	(4)
	OLS	First stage	Reduced form	2SLS
Dep. Var.	$\widehat{\text{FE}}_i^o$	$\log\left(\frac{\texttt{light}_i}{\texttt{light}_1}\right)$	$\widehat{\text{FE}}_i^o$	$\widehat{\text{FE}}_i^o$
$\log\left(\frac{\texttt{light}_i}{\texttt{light}_1}\right)$	–0.394***			–0.493***
	(0.061)			(0.079)
$\log\left(\frac{\overline{Price_i}}{\overline{Price_1}}\right)$		21.110***	–10.416***	
		(2.057)	(1.648)	
R-squared	0.268	0.552	0.228	
Observations	87	86	86	86
First-stage F-statistic				105

Note: This Table presents the OLS regression (column 1), the first stage (column 2), the reduced form version of the regression (column 3), and the IV estimate of Equation (24) (column 4). $\widehat{\text{FE}}_i^o$ are estimated using Equation (22). $\log\left(\frac{\texttt{light}_i}{\texttt{light}_1}\right)$ corresponds to the log of the nighttime light per capital normalized by the value associated to the reference group. $\log\left(\frac{\overline{\text{Price}_i}}{\overline{\text{Price}_1}}\right)$ corresponds to the log average world prices of the 5 most suitable crops normalized by the value of the reference group. The ethnic group Ifora is not included in columns 2 to 4 because the log ratio of prices is undefined for this group due to the underlying crop suitability being equal to zero. * Significant at 10%, ** significant at 5%, *** significant at 1%.

Figure D2: Estimated farming productivity $\widehat{A_i}$ and nighttime lights



Note: The x-axis displays the (logged) farming productivity $\widehat{A_i}$ recovered from the equation (29) for the 824 ethnic regions. The y-axis reports the (logged) nighttime lights per capita.

D.2 Material for step 5: Gravity of trade flows in Africa

In this section we estimate a gravity equation of bilateral trade between African countries over the 1970-2018 period to recover elasticities that are crucial for calibrating the inter-regional trade frictions τ_{in} (step 5 in the inversion procedure).

There is no data for trade flows at the detailed geographical level used for measuring bilateral violence. Our approach is therefore to apply coefficients obtained for bilateral trade equation estimated at the national level. We estimate the following country-level gravity equation using the data put together by Head and Mayer (2021) for the 1970-2018 period:

$$\mathbb{E}\left(X_{odt}\right) = \exp\left[\mu_1 \log \text{distance}_{od} + \mu_2 \text{Political border}_{od} + \mu_3 \text{RTA}_{odt} + \text{FE}_{ot} + \text{FE}_{dt} + \tilde{\epsilon}_{od}\right], \quad \text{(D.24)}$$

where X_{odt} is the bilateral trade flow from origin country o to destination country d in year t. The independent variables are the log of the bilateral distance between o and d, a binary variable that equals one if o and d are different countries, and another binary variable that equals one if o and d are part of the same regional trade agreement.[36] Estimation is carried out using the HDFE PPML estimator including origin×year and destination×year fixed effects, and clustering by dyad. Table D2 displays the estimates. The first column displays the traditional gravity setup that only considers international trade flows, for the whole panel. The coefficients on distance and RTA are of similar magnitudes when compared to the vast literature estimating those frictions (see Head and Mayer (2014) for a meta-analysis of those frictions). Column (2) focuses on intra-African trade, and shows that the absolute value of coefficient rises: distance is more detrimental to African trade, but RTAs have a larger positive influence. Those patterns are repeated in colums (3) and (4) which include self-trade for each of the countries and years. Crossing a national border reduces trade

[36] See the appendix of Head and Mayer (2021) for details of how flows internal to a country are constructed.

Table D2: Gravity estimates of frictions for Africa

Model	(1)	(2)	(3)	(4)	(5)	(6)
log distance (All)	-1.30***	-2.11***	-1.12***	-1.76***		
	(0.029)	(0.116)	(0.030)	(0.112)		
RTA (All)	0.349***	0.746***	0.856***	1.18***		
	(0.053)	(0.172)	(0.044)	(0.137)		
Border (All)			-3.45***	-4.21***		
			(0.080)	(0.186)		
log distance (not Afr)					-1.12***	
					(0.030)	
RTA (not Afr)					0.734***	0.504***
					(0.046)	(0.035)
Border (not Afr)					-3.14***	
					(0.084)	
log distance (Afr)					-0.845***	
					(0.038)	
RTA (Afr)					1.71***	1.07***
					(0.133)	(0.132)
Border (Afr)					-5.34***	
					(0.158)	
Observations	743,245	46,025	752,491	48,528	752,491	750,212
Squared Correlation	0.686	0.979	0.938	0.998	0.941	0.985
Only Africa:	No	Yes	No	Yes	No	No
With self-trade:	No	No	Yes	Yes	Yes	Yes

Note: All regressions using PPML. The first 5 columns include origin-year and destination-year fixed effects, while column (6) adds a country-pair fixed effect. Columns (2) and (4) restrict the sample to African countries, the others don't. Columns (1) and (2) do not include self-trade, the others do.

drastically in general, but even more in Africa. Another approach to allow for African specific frictions is to interact those 3 frictions with intra-African / not intra-African dummies. This allows for country-year fixed effects to be better estimated since all countries are included in the regression. The strong effects of African border crossing and RTAs is maintained, but distance effects are now more similar. Column (6) adds dyadic fixed effects to the regression letting only one variable to be identified in the within dimension: Again RTAs are revealed to have a stronger effect inside Africa. Results displayed in column 5 are used to calibrate the μs that are then used in equation (28).

D.3 Estimated parameters: summary

Table D3 presents descriptive statistics for the parameters estimated with our inversion procedure in Section 4. Note that this table reports estimates of s_n only for the 623 destination regions receiving violence; s_n is set to 1 for the remaining 201 regions which experienced no violence.

Table D3: Estimated parameters, some statistics

Param.	Numbers	Mean	Median	St. Dev.	Min	Max
Full sample						
$\widehat{A_i}$	824	3.970	3.885	1.541	0.394	13.595
Destinations						
$\widehat{s_n}$	623	0.992	0.998	0.024	0.628	0.99998
Origins						
$\widehat{\psi_i}$	87	0.014	0.011	0.012	0.0003	0.066
$\widehat{l_i}/pop_i$	87	0.078	0.018	0.159	0.001	0.864
$\widehat{L_i}/pop_i$	87	0.922	0.982	0.159	0.136	0.999

Note: Table D3 presents descriptive statistics for the parameters estimated with our inversion procedure in Section 4.

E Calibration of violence spillovers

E.1 Calibration of the destruction spillover

In the main text, equation (31) specifies the destruction spillover. In (31), the set of structural parameters to be calibrated consists of the semi-elasticity of TFP to violence ε_1 and of the vector of baseline productivity $\bar{\mathbf{A}}$. The calibration procedure thus consists in estimating the empirical counterpart of (31), such that:

$$\log A_n = -\varepsilon_1 \times \texttt{violence}_n + \text{residual}_{1,n}, \tag{E.25}$$

where A_n is recovered from step 6 of the model inversion procedure (page 22) and $\texttt{violence}_n$ represents the total violence as observed in the data.
An endogeneity concern arises from potential reverse causation going from productivity to violence: Indeed, a change in A_n affects local income per-capita w_n, which in turn can influence $\texttt{violence}_n$; this causal path is, in essence, linked to the *rapacity* effect. We address this issue by building a model-consistent shifter of $\texttt{violence}_n$. For this purpose, we incorporate the spatial

weights defined in equation (17) into equation (18) to express monadic violence exerted in n as:

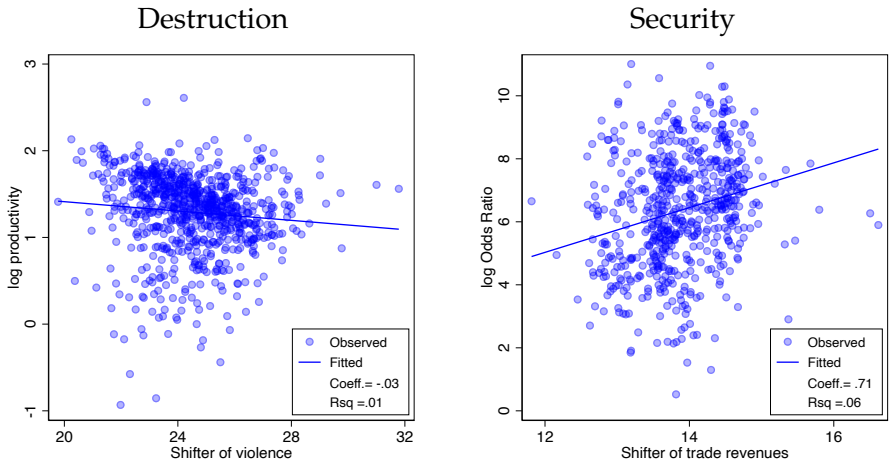
$$\log\left(\texttt{violence}_n\right) = \log(1-s_n) + \log w_n + \log \sum_i \frac{\xi_{in}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_i \bar{L}_i}{Y_i}\right)^{\frac{1}{1-\gamma}} \bar{L}_n}{\sum_k \xi_{kn}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_k \bar{L}_k}{Y_k}\right)^{\frac{\gamma}{1-\gamma}}}. \tag{E.26}$$

To satisfy the exclusion restriction, we need to choose a shifter of $\texttt{violence}_n$ that is independent of productivities and incomes. In this regard, the first two terms on the right-hand side of the previous equation are unsuitable, as they both depend on local income w_n, indirectly through the security spillover for the first term, and directly for the second term. A natural choice is the third term, which represents the SPV (see page 10) augmented with $\log \bar{L}_n$. After eliminating the dependence on incomes in this term, and using the observed population size, our model-consistent shifter of $\texttt{violence}_n$ is given by:

$$\widetilde{\text{SPV}}_n \equiv \log \sum_i \frac{\xi_{in}^{-\frac{\gamma}{1-\gamma}} (\psi_i \text{pop}_i)^{\frac{1}{1-\gamma}} \text{pop}_n}{\sum_k \left(\frac{\psi_k \text{pop}_k}{\xi_{kn}}\right)^{\frac{\gamma}{1-\gamma}}}. \tag{E.27}$$

We perform a 2SLS estimation of (E.25) using $\widetilde{\text{SPV}}_n$ as the instrument. The semi-elasticity of TFP to violence is estimated to be $-\varepsilon_1 = -0.00044$ with a standard error of 0.00017 and a first-stage F-statistic equal to 35. This elasticity has a direct quantitative interpretation: an increase of one hundred ACLED events leads to a reduction in local TFP by 4.4% . The corresponding reduced form relationship between the IV and the dependent variable is depicted in the left panel of Figure E3. Lastly, comparing equations (31) and (E.25) reveals that the structural interpretation of the error term is $\text{residual}_{1,n} = \log \bar{A}_n$. This relationship enables us to recover the baseline productivity level in the destruction spillover.

Figure E3: Reduced form evidence of violence spillovers



E.2 Calibration of the security spillover

We adopt a similar approach to calibrate the elasticity ε_2 and the baseline state capacity parameter $\bar{s}_n$ in the security spillover equation (32). This equation can be reformulated into the following

econometric specification:

$$\log \frac{s_n}{1-s_n} = \varepsilon_2 \times \log Y_n + \text{residual}_{2,n}, \tag{E.28}$$

where s_n is the state capacity parameter recovered from step 3 of the model inversion procedure (equation 26) and Y_n denotes the regional income as proxied by nighttime lights (see equation 24): $Y_n \equiv \texttt{light}_n^{\lambda} \times \texttt{pop}_n$.

To estimate this equation, we need to instrument income so as to address potential reverse causation from state capacity on income (through violence and the destruction spillover). We construct the following model-consistent shifter of n's trade revenues:

$$\widetilde{\texttt{SPT}}_n \equiv \log \sum_k \frac{\tau_{nk}^{1-\sigma} \bar{L}_k}{\sum_\ell \tau_{\ell k}^{1-\sigma}}, \tag{E.29}$$

which corresponds to a supply potential of trade based on equation (13), after removing all dependencies on income, productivity, and state capacity, and adapting the countries' subscripts such that n is the supplier of goods.

Equation (E.28) is estimated using $\widetilde{\texttt{SPT}}_n$ as the instrument. The 2SLS estimate of the elasticity of state capacity to income is equal to $\varepsilon_2 = 0.5$ with a standard error of 0.09 and a first-stage F-statistic equal to 284. The right panel of Figure E3 illustrates the reduced form relationship. Comparing equations (32) and (E.28) reveals that the structural interpretation of the residual is $\text{residual}_{2,n} = \log \frac{\bar{s}_n}{1-\bar{s}_n}$. This relationship allows us to recover the vector of baseline state capacity $\bar{\mathbf{s}}$.

F Pure Trade equilibrium

In the "Pure Trade" version of the model, the allocation of labor between production and fighting is held constant. This has consequences for the equilibrium characterization that differs in several respects from the system of equations reported in Section 2.3.

First, the labor market clearing condition (12) is still valid but L_i and l_i are now exogenously set. Second, in absence of free occupation choice, farming and fighting incomes—given by the two endogenous variables w_i^P and w_i^F, respectively—are not necessarily equalized in equilibrium and equation (10) is not verified anymore. Therefore, the wage dependence of the trade share in (13) and fighting probability in (14) must be adjusted accordingly. Finally, total income is now equal to $Y_i = w_i^F \times l_i + w_i^P \times L_i$. Combining these elements together, we obtain the set of equations that characterizes the pure trade equilibrium:

$$w_i^P L_i = \sum_{n=1}^{N} \frac{\tau_{in}^{-(\sigma-1)} \left(\frac{A_i}{w_i^P}\right)^{\sigma-1}}{\sum_{k=1}^{N} \tau_{kn}^{-(\sigma-1)} \left(\frac{A_k}{w_k^P}\right)^{\sigma-1}} s_n Y_n, \tag{F.30}$$

$$w_i^F l_i = \sum_{n=1}^{N} \frac{\xi_{in}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_i}{w_i^F}\right)^{\frac{\gamma}{1-\gamma}}}{\sum_{k=1}^{N} \xi_{kn}^{-\frac{\gamma}{1-\gamma}} \left(\frac{\psi_k}{w_k^F}\right)^{\frac{\gamma}{1-\gamma}}} (1-s_n) Y_n, \tag{F.31}$$

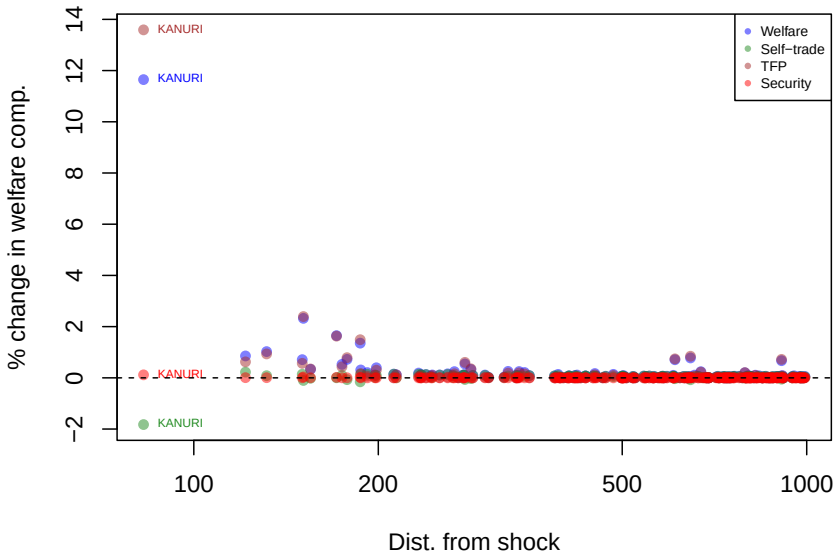
$$Y_i = w_i^F \times l_n + w_i^P \times L_i. \tag{F.32}$$

This system of $3 \times N$ equations has $3 \times N$ unknowns, represented by the vector $\{w_i^P, w_i^F, Y_i\}_{1 \leq i \leq N}$.

G Supplementary material for section 5 (counterfactuals)

G.1 Counterfactual 1: productivity shock

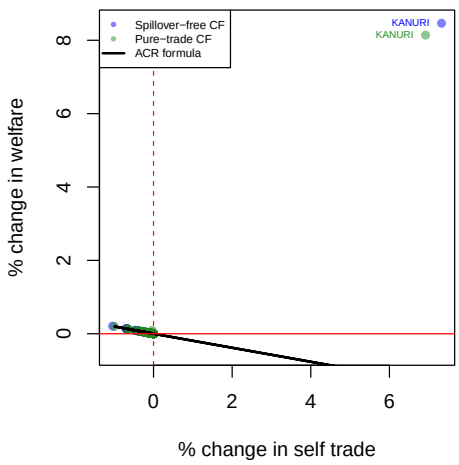
Figure G4: Welfare change decomposition after 10% rise of productivity in Kanuri's region



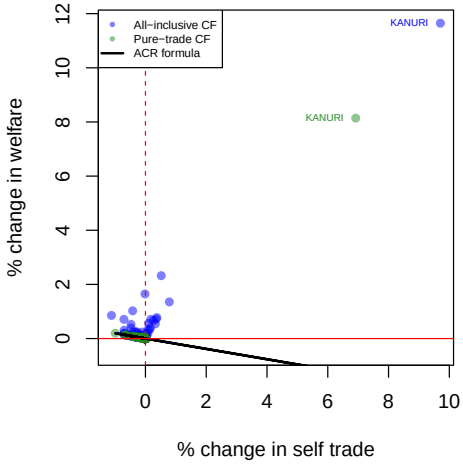
Note: The x-axis displays the distance from the 10% increase in productivity in the Kanuri region. The y-axis reports the overall welfare change in percentage points. All figures come from the "All-inclusive" scenario.

Figure G5: Welfare after 10% rise of productivity in Kanuri's region

(a) Welfare (ω_n, scenarios 1 & 2)



(b) Welfare (ω_n, scenarios 1 & 5)



Note: For all scenarios (1 and 2 in panel a; 1 and 5 in panel b), we compare the overall welfare gains in our model to the formula from Arkolakis et al. (2012). The x-axis displays the change in self trade (in percentage point). The y-axis reports the overall welfare change (in percentage points).

G.2 Counterfactual 2: Sahelxit

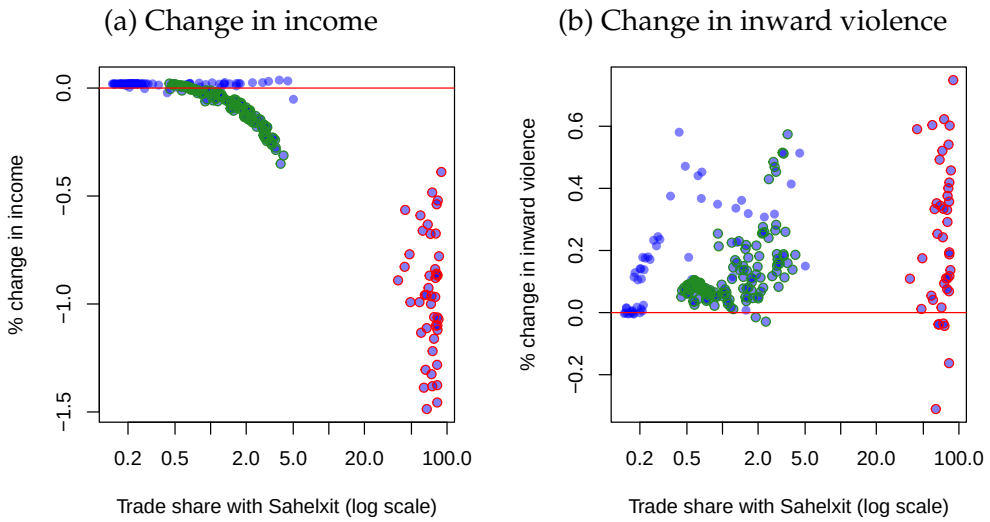
CEDEAO experiences significant levels of violence, with a total of 18543 events originating from the zone, the vast majority of which (17083) affect the region itself. Table G4 summarizes the bilateral flows of violence related to the zone. Most incidents are concentrated along the borders between Mali and Burkina Faso, as well as Mali and Niger. Notable hotspots of violence are also observed in Nigeria, Sierra Leone, and Liberia. The zone hosts two of the most violent origins in our sample: the Kanuri ethnic group, involved in 6166 events (4791 within CEDEAO) and the Masina ethnic group, responsible for 5300 events (all within CEDEAO).

Table G4: Summary of violence in CEDEAO region

Dest. → Origin ↓	Sahelxit	CEDEAO / Sahelexit	Rest of Africa	Total violence produced by origin
Sahelxit	6313 (97.29%)	105 (1.62%)	71 (1.09%)	6489 (100%)
CEDEAO / Sahelxit	1079 (8.95%)	9586 (79.53%)	1389 (11.52%)	12054 (100%)
Rest of Africa	462 (0.67%)	44 (0.06%)	68190 (99.26%)	68696 (100%)

Note: This Table presents a summary of the violence occurring in the CEDEAO region, considering three blocs of actors: armed groups located in the Sahelxit region (col. 2), armed groups located in the CEDEAO region after the Sahelxit (col. 3), and armed groups located in the rest of Africa (col. 4). Percentages represent the number of events from a specific origin×destination pair (e.g., Sahelxit actors attacking Sahelxit actors) relative to the total number of violent events produced by the considered origin (e.g., Sahelxit actors).

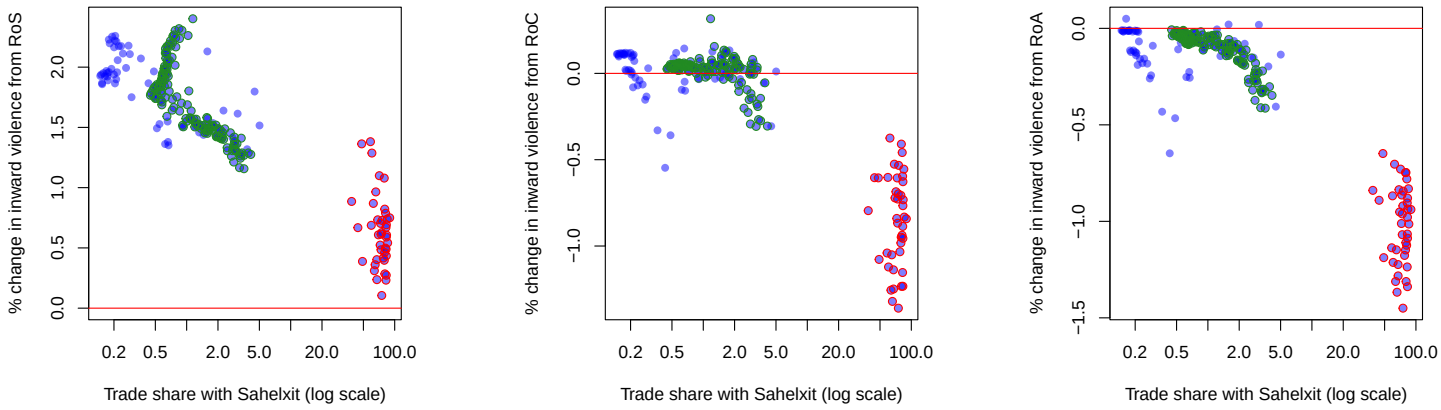
Figure G6: Income Loss and Violence after Sahelxit



Note: The x-axis displays the trade share with Sahelxit (log scale). In panel (a), the y-axis reports the % change in income. In panel (b), the y-axis reports the % change in inward violence.

Figure G7: Reshuffling of Violence after Sahelxit

(a) Violence from the Rest of Sahelxit (b) Violence from the Rest of CEDEAO (c) Violence from the Rest of Africa



Note: In all panels, the x-axis displays the trade share with Sahelxit (log scale). In panel (a), the y-axis reports the % change in inward violence from the Rest of Sahelxit regions. In panel (b), the y-axis reports the % change in inward violence from the Rest of CEDEAO regions. In panel (c), the y-axis reports the % change in inward violence from the Rest of Africa.

THE GRAVITY OF VIOLENCE*
– ONLINE APPENDIX –

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

September 17, 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 23th, 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.

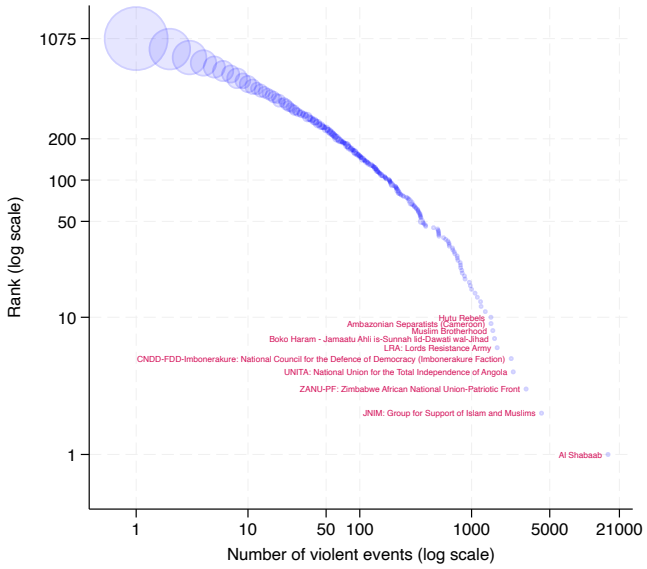
Step 4- Actor selection: In line with our theoretical model of raiding, we focus our analysis on violence perpetrated by actors classified as "Rebel groups" and "Political militias". These groups are indeed the most likely to project violence outside their rear base and to perpetrate violence-for-appropriation. 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.

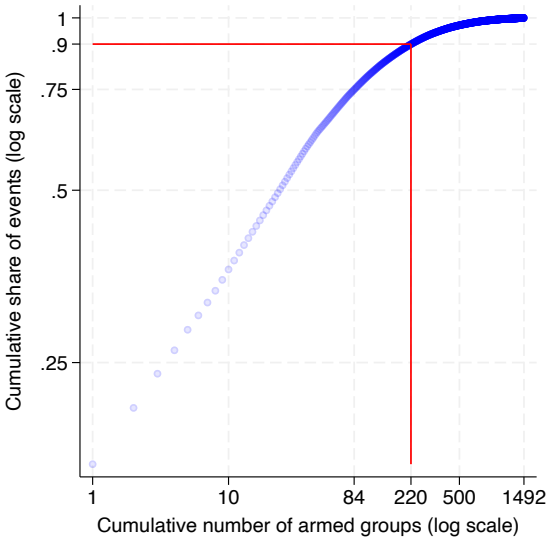
[8] At step 3, ACLED records a total of 553745 fatalities associated with 232878 events. Among those, rebel groups and political militias are responsible for 379314 fatalities in 87578 distinct events. Hence, the average number of fatalities per event for rebel groups and militias is 4.3. In contrast, 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]

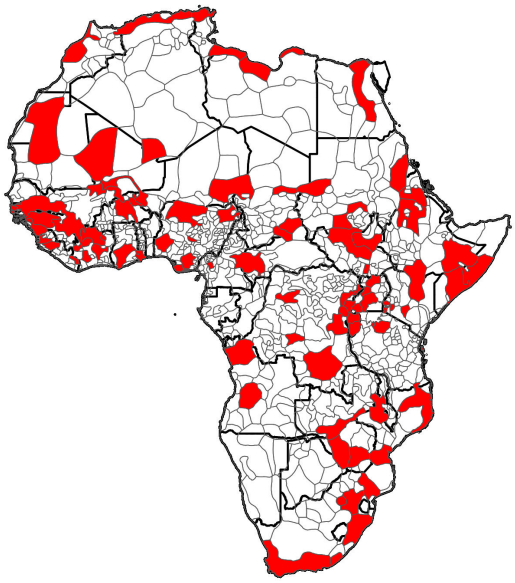
Step 5- Murdock filter on actors: To identify the origin of violence, we assign a unique Murdock ethnic group to the 220 violent actors. This task was conducted by the authors and independently cross-validated by a research assistant. The matching procedure involves three main steps: (5a) linking the 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 infor-

mation 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 156 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.

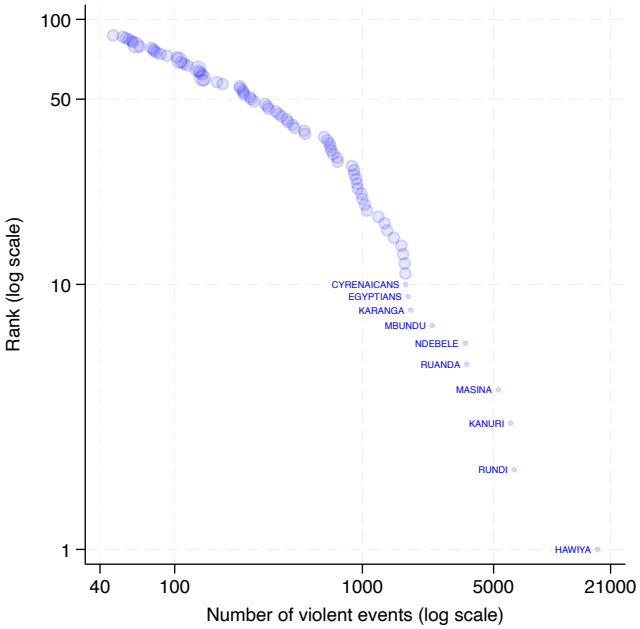
[25]The EPR Core dataset identifies all politically relevant ethnic groups and their access to state power in every country of the world from 1946 to 2021. It includes annual data on over 800 groups and codes the degree to which their representatives held executive-level state power—from total control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

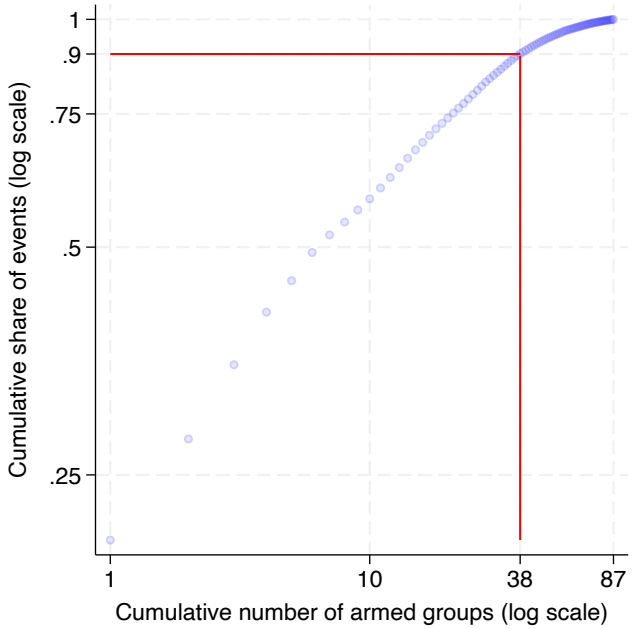
General Statistics: In Figure OA3.3 , we look at the distribution of violence originating from the 87 Murdock ethnic groups (step 6 of our data construction procedure). Panel (a) displays the log rank-log size relationship, highlighting the ten most violent ethnic groups, while panel (b) plots the cumulative 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 (15) 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:

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

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 $\lambda > 0$:

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

(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:

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

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:

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

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 $-\sigma < -1$):

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

Cancelling like terms:

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

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

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

implying that:

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

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:

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

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

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

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.