UNIVERSITÉ CATHOLIQUE DE LOUVAIN
FACULTY OF ECONOMIC, SOCIAL AND POLITICAL SCIENCES AND COMMUNICATION
ECONOMICS SCHOOL OF LOUVAIN
LOUVAIN INSTITUTE OF DATA ANALYSIS AND MODELING IN ECONOMICS AND STATISTICS
INSTITUTE OF ECONOMIC AND SOCIAL RESEARCH

Ethnicity, Conflict, and Identity

DOCTORAL DISSERTATION PRESENTED BY

DIEGO MALO RICO

IN FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR IN ECONOMICS AND MANAGEMENT

THESIS COMMITTEE:

Prof.	Joseph Flavian Gomes (Supervisor, Secretary)	UCLouvain
Prof.	Amma Panin (President, Chair)	UCLouvain
Prof.	Klaus Desmet (External member)	SMU
Prof.	Ómer Ozak (External member)	SMU
Prof.	Richard Bluhm (External member)	Stuttgart

LOUVAIN-LA-NEUVE, BELGIUM
AUGUST, 2025

Acknowledgements

I do not believe that success is solely the result of individual effort. Rather, it is shaped by the circumstances and the support of the people around us. In this sense, personal success is also a collective success. This thesis is a product of that collective effort, shared with the many people who have walked this path with me. Let me dedicate this section to them. Let's start from the beginning.

This thesis owes an immense debt to my parents, Diego and Paqui, for their unconditional support, just as important in the adjective as in the noun. Thank you for always answering the phone when I needed it, for your care, and for creating the circumstances that made this journey possible.

It is also indebted to the rest of my family: my sister, my grandparents, my uncles, my cousins, and the newest little ones. Building "Rico" has played a decisive role in my life, showing me what family means, and for being the place I call home. My family from Jaen is also a part of my identity, reminding me of where I come from and who I am. Each of you makes my life richer and more meaningful.

A part of who I am is also thanks to my friends. I am especially grateful to Jesús, for his time, his advice, and for always being there. Thanks as well to Mateo, Santi, Antonio, Luis, Roberto, Raúl, Emilio, and Facu. And to my neighbors, Franci and Alfonso.

The first step in my academic life began at the University of Malaga, with Professors Pablo, Ascensión, and Socorro. Alongside them, the Grupo Avanzado en Economía and the rest of the Economics faculty in Malaga guided and encouraged me at the very beginning of this journey.

The second step was taken at Universidad Carlos III de Madrid. I am doubly indebted here. On one hand, to my cohort, which was an endless source of support and inspiration. To my flatmates: Arnau, Henry, and Fernando; my office mates: Álvaro, Belén, Morteza (and Fernando again); and the rest of the cohort: Agustina, Florencia, Telmo, Lidia, Isabel, Iván, and Beka. On the other hand, I am grateful to the faculty and administrative staff at UC3M, who helped me grow both academically and personally. I want to especially mention Antonio Romero, Luisa Fuster, Luigi Minale, Carlos Velasco, and Antoine Loeper, who supported and guided me when I needed it most.

Before moving on to the third step, I want to mention TEVA, the company where I worked before moving to Belgium. I am especially grateful to Ricardo and Belen, my closest colleagues during that time.

The third step was possible thanks to Joseph, my supervisor. He gave me the opportunity to pursue a PhD, and I am deeply thankful for that. This thesis cannot be understood without his guidance, his patience, his understanding, and his way of being. He didn't just teach me how to conduct research or complete a PhD, he also showed me that how you do these things matters, and I hope to carry that lesson with me. If I ever have PhD students of my own, I hope to treat them the way he treated me.

This third step was also made possible by the Faculty of Economics at the Université Catholique de Louvain. From the very beginning, it was a place where I felt welcomed and comfortable, something essential to do a thesis. That was thanks to the faculty members: Jean-François, Gonzague, William, Vincent, Luca, David, François Maniquet, Michel, Bruno, Amma, Johannes, Muriel, and visiting professors like Akira; and to the administrative staff: Sébastien, Séverine, Flora, Virginie, and Manuella. And a big part of that welcoming environment was due to my fellow PhD students and postdocs, with whom I shared these years: Vincent, Keiti, Nathan, Guillermo, François, Angel, Chiara, Antoine, Gaia, Robin, Clara, Margaux, Francesco, Esmeralda, Silvia, Luigi, Charles, Ella, Thomas, Tiziano, Mathilde's, Marion, Yannik, Florian, Filippo, Marouane, Kam, and Rossana. I will always be proud to have done my PhD at IRES.

The friends I made in Belgium from the moment I arrived have also played a key role in this thesis. This work is also part of my office mate Rosa, of my close friend, Pablo, and of the friends who made me feel welcome from the very beginning and with whom I have shared beautiful moments: Matteo Bulgarelly, Matteo Maltonni, Laura, Rory, Carlos, Natalia, Valentina, Zhaniya, and Yahong.

Throughout this journey, there has been one person who has expanded my world in a way no one else could, who has taught me things that only love can teach. A person who reminded me of what matters, supported me through each stage, and made this journey deeply meaningful. For the space she opened in my life, for her care, for her love, for her presence, for walking with me. Thank you, Giulia.

This thesis has also benefited from the conversations and guidance of Klaus Desmet, with whom I was lucky to work. He provided me with invaluable insights on how to think, how to do research, and how to grow. I am also grateful for having him as one of my referees and in my PhD jury. Amma Panin has been an important part of this thesis from the very begining: since the discussant of my Master's thesis to later joining my PhD jury and being one of my referees. I have enjoyed our discussions, her courses, and her generosity. I also thank the other members of my PhD committee, Richard Bluhm and Ömer Özak, and a member of my comité d'accompagnement, François Maniquet. All of them have substantially improved the quality of this thesis and my future work. I am deeply grateful for their generosity, availability, and time. In addition, I want to thank the faculty at the University of Lausanne, and in particular Dominic Rohner, for hosting me during my research visit.

There are also people who have been important in my life and whom I have not yet mentioned: my friends from my first year at the Universidad de Jaén, my colleagues from the University of Malaga and the University of Lausanne, Claudia, Dario, Mané, Nicole, and Marionna. I have learned something valuable from each of you.

Finally, I want to thank the Universidad de Navarra and the Navarra Center for International Development for giving me the opportunity to continue my academic career.

I hope these words can, at least approximately, express how thankful I am. Thank you, everyone. Thank you for so much.

Contents

1

Introduction

This thesis is composed of three chapters that examine how ethnicity and global economic dynamics intersect to shape development outcomes in Sub-Saharan Africa, with a particular emphasis on conflict dynamics. It also explores how economic and cultural factors influence individuals' identification with different social groups.

In the first chapter, *"Ethnic Remoteness Reduces the Peace Dividend from Trade Access"*, co-authored with Prof. Klaus Desmet and Prof. Joseph Gomes, we examine whether a region's ethnic composition affects the extent to which it benefits from globalization. Specifically, we study the impact of a trade liberalization shock initiated by the United States, which reduced trade barriers for exports from Sub-Saharan African countries, thereby increasing international trade with the U.S. We find that regions more exposed to this shock experience less violent conflict and enjoy income gains. However, these positive effects are not observed in regions that are ethnically distant from the rest of their country. We argue that ethnic distance acts as a barrier to participating in the global economy, as these regions face difficulties integrating into national trade networks. We believe that our results should be considered in discussions about the winners and losers from globalization.

The second chapter, *"Beggar Thy Neighbor? Illicit Gold Trade and Conflict in the African Great Lakes Region"*, co-authored with my colleague Jing-Rong, investigates whether policies can exacerbate or mitigate conflict in resource-rich areas. We focus on the border region between Uganda and the Democratic Republic of Congo (DRC), analyzing the effects of a Ugandan tax exemption policy that led to the establishment of a gold refinery. Our findings show that after the refinery was established, violence increased in DRC regions with artisanal gold mining sites. We argue that the refinery boosted smuggling opportunities and increased the incentives for armed groups to control gold resources, thereby escalating violence against civilians. This represents a classic example of a "beggar-thy-neighbour" policy, where Uganda's domestic welfare improves at the expense of communities in the DRC.

In the third chapter, *"Culture or Economics? Social Identification in sub-Saharan Africa"*, I empirically test a theoretical economic model of social identification. The model posits that individuals choose group affiliations based on a trade-off: they prefer to identify with high-status groups, but if these groups are perceived as too socially or culturally distant, the cognitive cost of identifying with them increases. To quantify perceived distance, I construct a measure of linguistic distance between an individual's ethnic group and the national average. To measure status, I use the agricultural value of an individual's ethnic homeland. I find two key results: first, individuals from linguistically distant ethnic groups are more likely to identify with their ethnic group rather than the nation. Second, as the economic status of an ethnic group increases, its members are more likely to identify with their own ethnic groups, especially when the group is also linguistically remote. These findings offer empirical support for the model and contribute to our understanding of identity formation in heterogeneity societies.

Together, the results of these three chapters highlight the complex and often fragile interplay between identity, ethnicity, global integration, and economic incentives. They also provide a foundation for further research and offer valuable policy implications for both national governments and international stakeholders.

2

Ethnic Remoteness Reduces the Peace Dividend from Trade Access

This chapter is co-authored with Prof. Klaus Desmet and Prof. Joseph Gomes.

This paper shows that ethnically remote locations do not reap the full peace dividend from increased market access. Exploiting the staggered implementation of the Africa Growth and Opportunity Act (AGOA) and using high-resolution data on ethnic composition, violent conflict, and sectoral specialization for sub-Saharan Africa, we find that in the wake of improved trade access conflict declines less in locations that are ethnically remote from the rest of the country. We hypothesize that ethnic remoteness acts as a barrier that hampers participation in the global economy. Consistent with this, satellite-based luminosity data show that income gains from improved trade access are smaller in ethnically remote locations, and survey data indicate that ethnically more distant individuals do not benefit equally from positive income shocks when exposed to increased market access. These results underscore the importance of ethnic barriers when analyzing which locations and groups might be left behind by globalization.

2.1 Introduction

The starting point of this paper is three observations. First, positive terms of trade shocks affect the likelihood of conflict in developing countries. When such shocks increase the opportunity cost of conflict, they can potentially lead to a drop in violence, thereby resulting in a peace dividend.[1] Second, the gains from trade are limited not just by tariffs and transport costs, but also by other frictions, such as ethnic and linguistic barriers.[2] Third, ethnic differences are a fundamental driver of conflict around the world.[3]

Together, these observations raise the question of whether a location's ethnic composition might affect the potential peace dividend from improved trade access. Using high-resolution data for sub-Saharan Africa, this paper shows that after a positive trade access shock, there is an overall decline in conflict, but locations that are ethnically distant from the rest of the country benefit less from this peace dividend. In addition, such ethnically remote locations and ethnically remote individuals are more likely to be left behind by the income gains of globalization.

Exploiting geographic and temporal variation in trade access across sub-Saharan Africa, we explore how a location's ethnic remoteness mediates the impact of improved market access on conflict. Our premise is that a location's ethnic remoteness, defined as its population-weighted average ethnic distance to the rest of the country, acts as a barrier to accessing local trade networks and power structures that facilitate integration into the global market.[4] To get

[1] Berman and Couttenier (2015) provide evidence of positive terms of trade shocks lowering conflict in sub-Saharan Africa, whereas Dix-Carneiro et al. (2018) show how negative terms of trade shocks increase crime in Brazil. Dube and Vargas (2013) present a more mixed picture, arguing that the benign effect of positive terms of trade shocks on conflict is limited to commodities that are labor-intensive.

[2] For evidence on ethnic and linguistic barriers to trade, see Isphording and Otten (2013), Melitz and Toubal (2014) and Aker et al. (2014).

[3] Papers that have studied the link between ethnicity and conflict include Fearon and Laitin (2003), Collier and Hoeffler (2004), Montalvo and Reynal-Querol (2005), Esteban et al. (2012a) and Esteban et al. (2012b).

[4] In the baseline, we focus on the average distance when defining ethnic remoteness, because in sub-Saharan Africa power tends to be assigned proportionally to the sizes of ethnic groups (Francois et al., 2015). As a robustness check, we also use an alternative measure, based on the population-weighted average ethnic distance to the country's largest ethnic group.

temporal and spatial variation in trade access, we rely on the Africa Growth and Opportunity Act (AGOA), which during the 2000s lowered U.S. trade barriers for most African countries. Because not all African countries were part of AGOA, and because accession occurred in a staggered manner, there is cross-country and cross-time variation in trade access. By further interacting country-level exposure to AGOA with within-country geographic variation in proximity to the closest port and in AGOA eligibility of local production, we also exploit within-country local variation in trade access. Combining the sub-national trade access data with high-resolution geo-coded data on ethnic remoteness and conflict, we can analyze how the effect of trade liberalization on conflict depends on a location's ethnic remoteness.

At a spatial resolution of 55 Km $\times$ 55 Km, we regress the intensity of conflict on local exposure to AGOA and on the interaction of this exposure with ethnic remoteness. Identification relies on including grid-cell fixed effects as well as country-time fixed effects in our empirical specification. These fixed effects purge estimates of time-invarying cell-level and time-varying country-level unobservable characteristics that might pose a threat to causality. For example, accession to AGOA depended partly on a country's democratic freedoms and its respect for private property rights, but these characteristics are also likely to affect conflict. Country-time fixed effects absorb any such impact. In addition to including fixed effects, we control for time-varying cell-level weather shock variables that have been found to be important for conflict (Burke et al., 2015), and for a wide range of potentially confounding cell-level variables interacted with local exposure to AGOA.

Our cell-level regressions establish two main results. First, locations that experience greater improvements in market access suffer less from violent conflict: accession to AGOA lowers conflict, and more so in locations that are closer to ports. There is thus a peace dividend from trade access. Second, being in an ethnically more remote location mitigates this positive effect. That is, the benefits of accession to AGOA on conflict are partly or wholly wiped out in locations that are ethnically distant from the rest of the country. This latter result is not driven by ethnically remote locations also being geographically remote.

These findings are robust to alternative ways of measuring exposure to AGOA. In the baseline, we define a cell's exposure to AGOA as its proximity to the nearest port, conditional on the country being part of AGOA and on the cell producing AGOA-eligible goods. As a first alternative, we consider a broader definition of exposure that does not condition on a cell producing AGOA-eligible goods. In that case, within-country spatial variation in exposure comes only from differences in proximity to the nearest port. As a second alternative, we consider a narrower definition of exposure that conditions our baseline measure on a cell producing AGOA-eligible goods in which the country already had export capacity in the pre-AGOA period. As a last alternative, we condition exposure on the land suitability of cells for AGOA-eligible crops, rather than on the actual production of such crops. When using any of these alternative exposure measures, our results are unchanged.

In addition to ethnic remoteness, a location's ethnic composition might mediate the relation between market access and conflict in other ways. In particular, a location's ethnic diversity and its ethnic complementarity might matter too. A location's ethnic diversity measures to what extent its ethnic groups are fractionalized (Easterly and Levine, 1997; Alesina et al., 2003) or polarized (Esteban et al., 2012a; Montalvo and Reynal-Querol, 2005). Ethnically diverse places typically find it harder to build consensus and reach agreements. When faced with an increase in contestable income in the wake of a positive trade shock, we might therefore expect ethnically diverse locations to resort to violence (Fearon and Laitin, 2003; Collier and Hoeffler, 2004; Montalvo and Reynal-Querol, 2005). Our paper finds no robust evidence of this mechanism. A location's ethnic complementarity, for its part, measures to what extent its ethnic groups depend on each other. Greater interdependence might facilitate sharing the gains from trade, so we might expect ethnic complementarity to reduce conflict (Jha, 2013). Our paper finds no empirical support for this mechanism either. Instead, only a location's ethnic remoteness affects the peace dividend from trade access. Controlling for additional measures of ethnic interdependence such as kinship tightness and segmentary lineage does not affect these results (Enke, 2019; Moscona et al., 2020).

What mechanism might explain our findings? Trade theory predicts that easier access to foreign markets through AGOA should imply income gains from trade. However, the relation between higher income and conflict is not without ambiguity. On the one hand, the opportunity cost effect emphasizes that positive income shocks make it more costly to engage in conflict. On the other hand, the rapacity effect emphasizes that positive income shocks increase contestable income, giving rise to more conflict (Dube and Vargas, 2013; Bazzi and Blattman, 2014; Berman et al., 2017; Blair et al., 2021).[5] Our finding of a peace dividend from AGOA is consistent with the opportunity cost effect, rather than with the rapacity effect. Of course, improved market access through AGOA does not do away with all trade costs. There continue to be trade frictions in the form of transport costs, linguistic barriers, and more generally, any other friction that limits effective integration into the world market. To the extent that ethnically remote locations face greater frictions to access the world market, we would expect them to benefit less from the positive effect of trade liberalization on conflict. This is consistent with our finding of a reduced peace dividend from AGOA in ethnically remote locations.

This interpretation of our results relies on AGOA having a positive income effect that is weakened by ethnic remoteness. However, so far, we have not provided any evidence of the effect of AGOA on income. We therefore investigate whether cells that are more exposed to AGOA experience greater income gains as proxied by increases in nighttime luminosity, and whether cells that are ethnically more remote experience smaller gains. We use the exact same empirical specification as before, with the difference that we now look at the effect of the AGOA trade shock on luminosity rather than on conflict. Consistent with our interpretation, we find that accession to AGOA increases luminosity more in cells that are exposed to the AGOA shock, though this positive effect is smaller in cells that are ethnically more remote.

As further evidence for this income effect, we also use individual-level data from the different waves of the Afrobarometer. We find that individuals that are ethnically more distant from the rest of the country suffer negative income shocks when exposed to increased trade, compared to individuals that are ethnically less distant. When estimating this effect, we are able to control for a wide range of individual characteristics, such as age, gender, ethnicity and profession. Including profession purges estimates of possible effects coming from differences in specialization, and including ethnicity allows us to control for any effect of within-group genetic diversity (Arbatlı et al., 2020).

An important contribution of our paper is the construction of a 0.5×0.5 dataset for sub-Saharan Africa on production across all sectors of the economy. Existing research on trade shocks and conflict primarily relies on international price fluctuations as exogenous variation, limiting analysis to a narrow set of commodities with accessible price data (e.g., Armand et al.,

[5] In contrast to our work, these empirical studies do not address the possible role of ethnic composition. For a theoretical analysis of these two effects, see Dal Bó and Dal Bó (2011).

2020; Berman et al., 2017; McGuirk and Burke, 2020; Dube and Vargas, 2013). In contrast, our method maps AGOA-eligible tariff lines to a comprehensive range of commodities, allowing us to move beyond specific crops or minerals and achieve a broader, more precise identification of trade shocks. Further, while existing studies often focus on either agriculture (e.g., Berman and Couttenier, 2015) or mining (e.g., Berman et al., 2017), our approach not only includes a highly granular set of crops, minerals, and oil fields but also expands to an additional 93 industries spanning manufacturing, textiles, and apparel. This enables us to construct a unique dataset with approximately 40,000 observations for these industries, geolocated across 7,142 sites in sub-Saharan Africa. Starting with NAICS-level industry data, we map each NAICS code to specific U.S. tariff lines, incorporating both AGOA and GSP provisions to enhance the dataset's granularity and utility. We expect this dataset to have value beyond the scope of our paper.

Our paper is related to a large literature on the effect of terms of trade shocks on conflict. Closest to our work is Berman and Couttenier (2015) who show that positive terms of trade shocks in sub-Saharan Africa lower conflict, but less so in geographically more remote places. However, they do not explore the relation between trade liberalization, ethnicity and conflict, which is the focus of this paper. Other work that analyzes the relation between trade and conflict also ignores the ethnic dimension (Barbieri and Reuveny, 2005; Dix-Carneiro et al., 2018; Martin et al., 2008a,b, 2012).

Our interest in ethnic remoteness draws on the trade literature that has explored the role of linguistic and ethnic barriers as additional trade frictions (Isphording and Otten, 2013). These costs are not simply related to having a common language. Ethnic ties matter beyond their effect on the ease of communication (Melitz and Toubal, 2014). Trade frictions do not only exist between countries, but they also exist within countries. For goods to be shipped overseas, they first need to successfully get to a port. This involves not just overcoming within-country geographic barriers but also within-country ethnic barriers. As an illustration, Aker et al. (2014) find within-country ethnic borders in Niger to be comparable to national borders in how they limit trade.[6]

Ethnic, linguistic or genetic distances have also been shown to matter for other outcomes, such as human capital accumulation (Laitin and Ramachandran, 2016; Shastry, 2012), labor market outcomes of immigrants

[6] In related work, Boken et al. (2023) document that in West Bengal caste differences act as barriers to firm-to-firm trade.

(Isphording, 2014), the diffusion of ideas (Spolaore and Wacziarg, 2009), market integration (Fenske and Kala, 2021), and the effectiveness of counterinsurgency policies (Armand et al., 2020). Recent work has taken a more micro approach, using high-resolution geographic data or individual-level data to study ethnic barriers. For instance, Gomes (2020) highlights how ethnic distance to neighbors impedes access to health information, leading to higher child mortality in sub-Saharan Africa. By linking ethnic remoteness to local conflict outcomes in the context of trade shocks, we add a new dimension to understanding how ethnic barriers shape the instability-inducing effects of globalization.

Our paper speaks to the question which groups and locations are left behind by globalization. The differential impact of trade liberalization on skilled and unskilled workers is a well-studied phenomenon (Goldberg and Pavcnik, 2007). More recent work has turned its focus to geography, comparing regions that are differentially affected by either lower import tariffs or improved market access. For example, Topalova (2010) finds smaller declines in poverty in Indian districts that experienced greater tariff reductions in the wake of India's 1991 trade liberalization, whereas McCaig (2011) finds faster declines in poverty in Vietnamese provinces that benefited more from improved market access after the signing of the U.S-Vietnam Bilateral Trade Agreement in 2001. In developed countries, the so-called China trade shock has drawn much attention. Areas in the U.S. that were more exposed to Chinese import competition experienced deteriorating economic conditions (David et al., 2013; Autor et al., 2014, 2016). In these different studies of who might benefit and who might be left behind by globalization, the ethnic dimension has been ignored.[7] We find that both ethnically remote locations and ethnically remote individuals fail to reap the full benefits of improved trade access.

2.2 Data

Using a 0.5×0.5 spatial grid (approximately 55 km by 55 km at the Equator), this paper empirically analyzes how ethnic remoteness mediates the effect of

[7]This is a major omission as inequality between ethnic groups can have severe pernicious effects on both economic growth (Alesina et al., 2016) and violent conflict (Mitra and Ray, 2014).

trade access on conflict.[8] We also consider how ethnic diversity and ethnic complementarity might act as separate channels affecting the relation between trade access and conflict. The time frame of our study goes from 1989 to 2017, and we focus on sub-Saharan Africa.

By combining time-varying country-level accession to the Africa Growth and Opportunity Act (AGOA) with within-country variation in proximity to the closest port and in the production of AGOA-eligible goods, we construct a measure of trade access that varies across time and space. To measure ethnic remoteness at the cell level, we rely on high-resolution data on the location and size of ethnolinguistic groups. By using geo-coded data on conflict from UCDP and ACLED, we explore how ethnic remoteness affects the peace dividend from increased market access. Proxying local income by nighttime light intensity, we also analyze whether ethnic remoteness acts as a barrier that limits the income gains from trade. The rest of this section describes the data in more detail. Appendix 2.A.1 provides a detailed list of data sources, and Appendix Tables B1 and B2 report summary statistics and cross-correlations of the main variables of interest.

2.2.1 Dependent Variable: Conflict or Income

Conflict. As main source for our geo-coded conflict data, we use the UCDP Georeferenced Event Dataset, covering all 48 sub-Saharan African countries in our study for the period 1989–2017. This dataset defines violence as the use of "armed force by an organized actor against another organized actor or against civilians" (Sundberg and Melander, 2013, p. 524). Organized actors include governments of independent states or non-governmental organized groups. For our study, we aggregate the conflict data up to the 0.5×0.5 grid-cell level.

As an alternative, we also use the Armed Conflict Location and Event Data (ACLED). This dataset takes a broader view of political violence by including civil and communal conflicts, violence against civilians, and rioting and protesting. One disadvantage of ACLED is that it starts in 1997, only three years before the enactment of ACLED. That makes the longer time span of UCDP somewhat more attractive for our purpose. However, we conduct

[8]The 0.5×0.5 spatial grid based on PRIO has been used extensively in the literature. See, for instance, McGuirk and Burke (2020), Berman and Couttenier (2015), and Berman et al. (2017). Cells that overlap the borders of two or more countries are split into smaller sub-cells pertaining to distinct countries.

extensive robustness analysis using the ACLED data.[9]

Income. Following the pioneering work by Henderson et al. (2012), a large number of papers have used nightlight as measured by satellites as a proxy for income.[10] For 1992–2013 we use the DMSP-OLS Nighttime Lights Time Series v.4, whereas for 2013–2017 we use the DMSP-like Nighttime Lights Derived from VNL, an extend series of annual nighttime lights using VIRSS data developed by Nechaev et al. (2021). This gives us a cell-level panel dataset of luminosity for 1992–2017. Intensity of luminosity, coded at the grid-cell level, takes values ranging from 0 (no lights) to 63 (maximum luminosity).

2.2.2 Trade Access

To identify the effect of market access, we rely on the Africa Growth and Opportunity Act of 2000 that gave sub-Saharan African countries preferential trade access to the United States.

Trade access through AGOA. Because not all countries became part of AGOA and because accession occurred in a staggered manner, there is cross-time and cross-country variation in trade access. To get within-country variation in trade access, we rely on two sources of local variation: proximity to major ports,[11] and production of AGOA-eligible goods in the pre-AGOA period. The closer a location is to a major port, the more it gains market access when joining AGOA. However, market access only improves if the location already produced AGOA-eligible goods.

By multiplying country-level trade access by a cell-level measure of proximity to the nearest port and a cell-level binary measure of production of AGOA-eligible goods, we get a cell-level time-variant measure of trade access:

$$AGOAccess_{ict} = AGOA_{ct} \times Proximity_{ic} \times \max_{j \in J}\{Production_{icj}\} \tag{2.1}$$

[9]For other papers that use UCDP and/or ACLED, see Berman and Couttenier (2015), McGuirk and Burke (2020), Armand et al. (2020), Cervellati et al. (2022), and Moscona et al. (2020).

[10]See Michalopoulos and Papaioannou (2018) for a review of this literature.

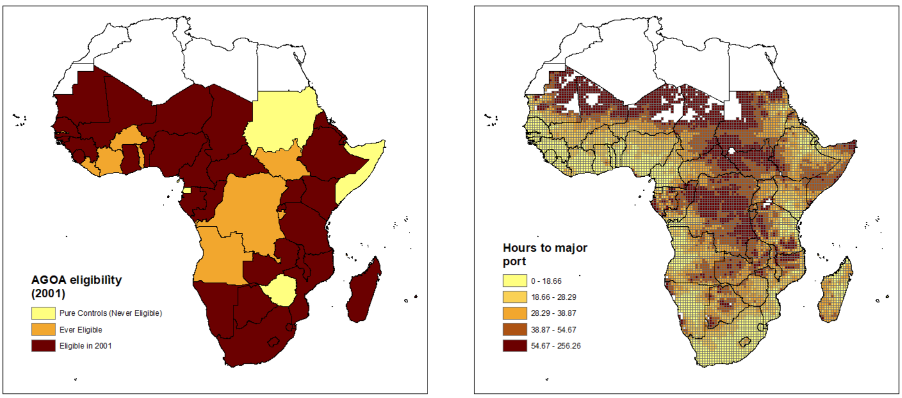
[11]Using proximity to a major port is reasonable, since more than 90% of international trade in Africa relies on maritime transport (Sebastian, 2014).

where $AGOAccess_{ict}$ denotes trade access in cell i of country c in year t, $AGOA_{ct}$ denotes whether or not country c was part of AGOA in the year t, $Proximity_{ic}$ denotes the proximity of cell i of country c to the nearest major port in 2000, and $Production_{icj}$ denotes whether or not cell i of country c produced good j in the pre-AGOA period, where J is the set of AGOA-eligible products.

To get a measure of $Proximity_{ic}$, we standardize the number of hours required to travel to the nearest major port from IFPRI, and subtract this standardized variable from its maximum. Part (a) of Figure 2.1 maps the cross-country variation in access to AGOA, whereas part (b) of Figure 2.1 depicts the travel time to the nearest major port expressed in hours.

Figure 2.1. Trade Access through AGOA

(a) AGOA Membership (b) Travel Time to Nearest Port



Notes: Panel (a) plots three types of countries: (i) countries that could have entered AGOA but never did (pure controls); (ii) countries that entered AGOA for at least one year during the period of our study; (iii) countries that were eligible for AGOA in 2001, i.e., the first year of its implementation. The North African countries in white were never part of AGOA and are not part of our sample. Panel (b) plots the travel time to the nearest major port in hours for the year 2000 (i.e., pre-AGOA). A higher travel time to port represents a lower degree of trade openness, as approximately 90% of African trade is maritime.

To measure $Production_{icj}$, we determine whether a cell i in country c produces AGOA-eligible product j. Constructing this high-resolution sectoral database for sub-Saharan Africa is a key contribution of this paper. More specifically, we match the tariff lines of all products included in AGOA to geolocated data on the production of oil and 19 AGOA-eligible minerals (from a list of 33), 72 AGOA-eligible crops (from a list of 175), and 93 AGOA-eligible textile, apparel, agricultural and other manufacturing industries. The tariff lines are for the year 2000, and are based on publicly available data from the United States International Trade Commission (USITC).[12] Locations of oil fields and mines come from PETRODATA and SNL Metals & Mining dataset (S&P Global Marketplace); crop locations are based on Ramankutty et al. (2008); and locations of textile, apparel and other manufacturing sectors at the 4-digit NAICS level are from Dunn & Bradstreet. Appendices 2.A and 2.C provide further details on the construction of these data.

Figure 2.2 plots the cells that produce different categories of AGOA-eligible products. We consider a cell as treated if it produces AGOA-eligible crops, minerals, or oil, or if it houses manufacturing units producing AGOA-eligible products, provided the country is eligible for AGOA in the particular year. Additionally, we also consider a cell as treated if it hosts textile or apparel production units, provided the country is eligible for apparel or textile provisions under AGOA in that particular year.[13] Using this definition, Figure 2.3a depicts all cells that were treated in at least one year.

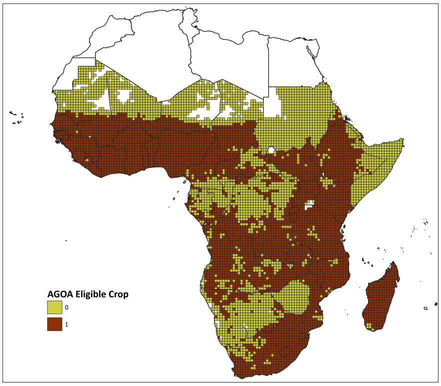
Accession to AGOA depended mostly on countries having some basic level of private property rights, rule of law, democratic freedoms, and a market-based economy.[14] Differences in such rights, freedoms, and institutions partly explain why some countries, such as Somalia, never became eligible, why other countries, such as Sierra Leone, were admitted late, and why a few countries, such as Eritrea, were removed. Appendix Table A1 lists the full list of countries that were ever eligible for AGOA along with years of eligibility, and Appendix Table A2 does the same for the textile and apparel sectors. To the extent that accession criteria are related to conflict, we might face an endogeneity problem. We address this potential issue by including country $\times$ year fixed effects in all our regressions.

[12] In the specific case of apparel, textiles, and other manufacturing sectors, each 4-digit NAICS industry tends to match to many different tariff lines. In that case, we consider a sector to be treated by AGOA if at least 25% of the tariff lines are AGOA-eligible.

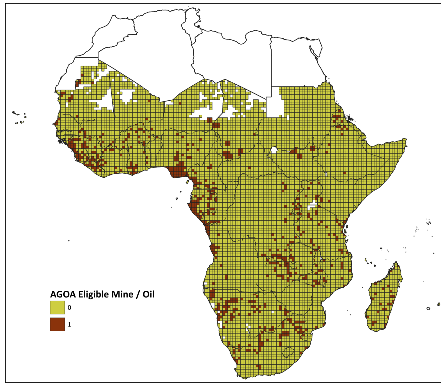
[13] AGOA eligibility for textiles and apparel are different from general AGOA eligibility. For example, during the time frame of our study the AGOA membership of Guinea did not include apparel.

[14] See https://agoa.info/about-agoa/country-eligibility.html.

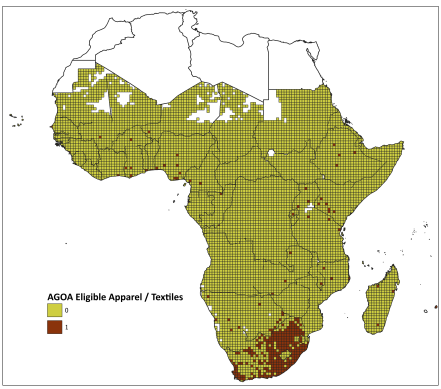
Figure 2.2. Location of AGOA-Eligible Products



(a) AGOA-Eligible Crops



(b) AGOA-Eligible Minerals and Oil



(c) AGOA-Eligible Apparel and Textiles

(d) AGOA-Eligible Manufactured Goods

Notes: Panel (a) plots all cells that produce AGOA-eligible crops. Panel (b) shows cells producing AGOA-eligible minerals or oil. Panel (c) shows those producing AGOA-eligible apparel and textiles. Panel (d) shows cells with AGOA-eligible manufactured goods. See Appendix 2.A.1 for sources and definitions.

Of course, to use AGOA as a shock to trade access, ideally it needs to have a sufficiently large effect on exports. Focusing on the program's three key product categories (apparel, agriculture, and manufactures), Frazer and Van Biesebroeck (2010) estimate an AGOA-induced increase in exports of 34%. Looking more broadly at all non-oil exports, the effect was a more modest, but still not trivial, 8.0%.

Alternative measures of trade access through AGOA. For robustness, we consider three further measures of time-varying cell-level exposure to AGOA. A first alternative measure defines exposure to AGOA more broadly than our baseline measure (2.1):

$$AGOAGeo_{ict} = AGOA_{ct} \times Proximity_{ic} \tag{2.2}$$

In this case, a location's market access depends on proximity to the nearest port, but not on it producing AGOA-eligible goods in the pre-AGOA period. It aims to capture the idea that areas closer to the port experience a general improvement in trade access, regardless of their production structure.

A second alternative measure defines exposure to AGOA more narrowly:

$$AGOAExp_{ict} = AGOA_{ct} \times Proximity_{ic} \times \max_{j \in J}\{Export_{cj} | Production_{icj} = 1\} \tag{2.3}$$

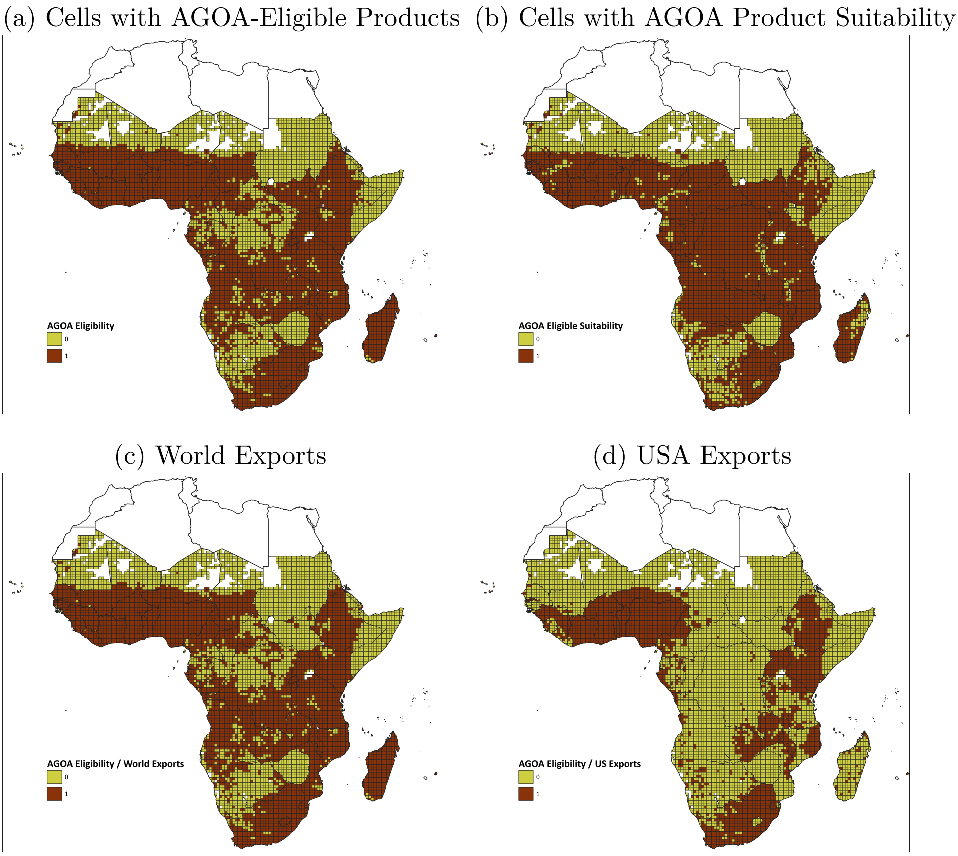
where $Export_{cj}$ is a binary variable that indicates whether country c exported good j in the pre-AGOA period. To measure export capacity, we use data from CEPII. We set two different bars for a country's export capacity in a certain good by requiring positive exports to either anywhere in the world (see Figure 2.3c) or the U.S. (see Figure 2.3d). Exposure measure (2.3) takes the view that if a location produces AGOA-eligible goods, but the country has no export capacity in those goods, then the cell will not experience an improvement in market access when the country joins AGOA.

A third alternative measure uses land suitability to define crop exposure to AGOA:

$$AGOASuit_{ict} = AGOA_{ct} \times Proximity_{ic} \times \max_{j \in J}\{Suitability_{icj}\} \tag{2.4}$$

where $Suitability_{icj}$ measures whether a location's land is suitable for AGOA-eligible crop j using data from the FAO's Global Agro-Ecological Zones (GAEZ) database. For all other AGOA-eligible products (such as minerals, oil, textiles and apparel), we continue to use actual production. Exposure measure 2.4 takes the view that as long as the land is suitable for the production of AGOA-eligible goods, the location experiences an improvement in market access when joining AGOA (see Figure 2.3b) .

Figure 2.3. Trade Access through AGOA: Alternative Definitions



Notes: Panel (a) plots all cells that have AGOA-eligible products. Panel (b) plots all cells that have adequate conditions making them suitable for producing AGOA-eligible products. Panel (c) plots all cells that have AGOA-eligible products conditional on the country exporting that product to anywhere in the world in the pre-AGOA period. Panel (d) plots all cells that have AGOA-eligible products conditional on the country exporting that product to the U.S. in the pre-AGOA period. See Appendix 2.A.1 for data sources and variable definitions.

2.2.3 Ethnic Remoteness

In sub-Saharan Africa ethnicity and language largely overlap. Data on the population's ethnic composition at the 0.05 degree $\times$ 0.05 degree grid-cell level come from the language database recently constructed by Desmet et al. (2020). They combine three sources of information: data on the spatial distribution of population from Landscan, data on the linguistic composition of countries from Ethnologue (Lewis et al., 2014), and maps on the geographic distribution of 6,905 distinct languages from the World Language Mapping System (WLMS). Using this information, they implement an iterative proportional fitting algorithm to construct a comprehensive 0.05 degree $\times$ 0.05 grid-cell level dataset on the ethnolinguistic composition of the population for the entire globe. We aggregate this information up to the 0.5×0.5 grid-cell level.

Ethnic remoteness aims to proxy for the ethnic barriers that residents of a location face in accessing local trade networks and power structures that facilitate their integration into the global market. When measuring ethnic remoteness, we can either take remoteness to the country or remoteness to the dominant group. In the context of sub-Saharan Africa, Francois et al. (2015) find a high degree of proportionality in the assignment of power positions between ethnic groups. As main measure of a cell's ethnic remoteness, we therefore take the average ethnic distance between a random resident of the cell and a random resident of the country. To be more precise, consider a country partitioned into different grid-cells indexed by ℓ or k with a population belonging to different ethnic groups indexed by n or m. Denote by d_{nm} the ethnic distance between n and m, by s_n the share of the country's population pertaining to ethnic group n, and by $s_{\ell n}$ the share of the population of grid-cell ℓ pertaining to ethnic group n. We then define the ethnic remoteness of cell ℓ to the country as

$$ER_\ell = \sum_n \sum_m s_{\ell n} s_m d_{nm}. \tag{2.5}$$

Given that in Africa ethnicity tends to coincide with language, we measure d_{nm} as the linguistic distance between the language spoken by ethnic group n and the language spoken by ethnic group m (Gomes, 2020). Following a large literature, we use a linguistic distance measure that is based on the number of shared branches in a linguistic tree.[15] More specifically, we take the Ethnologue language tree, and denote by b_{nm} the number of shared branches between languages n and m, and by b_{max} the maximum number of shared branches between any two languages. We then define the linguistic distance between n

[15]See, for instance, Fearon (2003a), Desmet et al. (2009a), Desmet et al. (2012), Esteban et al. (2012a), Esteban et al. (2012b), Laitin and Ramachandran (2016) and Gomes (2020) for a similar approach.

and m as

$$d_{nm} = 1 - \left(\frac{b_{nm}}{b_{max}}\right)^{\delta} \tag{2.6}$$

where δ is a parameter that determines how fast the linguistic distance declines as the number of shared branches increases. We follow Desmet et al. (2009a) and set $\delta = 0.05$.

Panel (a) of Figure 2.4 shows a grid-cell map of ethnic remoteness in sub-Saharan Africa. One relevant question is to what extent ethnic remoteness is distinct from geographic remoteness. The correlation between ethnic remoteness and travel time to the nearest port is only 0.255. This clarifies that ethnic remoteness captures a concept that is distinct from geographic remoteness.

As an alternative measure to ethnic remoteness to the country average, we also consider the ethnic remoteness of a cell ℓ to the country's dominant group:

$$ER_{\ell}^{dom} = \sum_{n} s_{\ell n} d_{n,dom}, \tag{2.7}$$

where $d_{n,dom}$ is the distance between ethnic group n and the country's largest ethnic group dom.

2.2.4 Ethnic Diversity

Although our main focus is on ethnic remoteness, we also consider whether other aspects of a location's ethnic composition might mediate the relation between trade and conflict. It has been widely documented that ethnic diversity is a fundamental driver of conflict in sub-Saharan Africa (Collier and Hoeffler, 2004). We consider two measures of a cell's ethnic diversity. One is the standard fractionalization index, which measures the probability that two randomly drawn individuals of cell ℓ pertain to different ethnic groups:

$$ELF_{\ell} = \sum_{n} \sum_{m} s_{\ell n} s_{\ell m}. \tag{2.8}$$

Another is the standard polarization index from Montalvo and Reynal-Querol (2005), which measures the proximity of the distribution of the populations of ethnic groups in a cell to a bipolar distribution (i.e., a distribution with two ethnic groups each having a population of 50%):

$$POL_{\ell} = \sum_{n} s_{\ell n}^{2}(1 - s_{\ell n}). \tag{2.9}$$

In the robustness checks, we will also consider other fractionalization and polarization indices that take into account distances between ethnic groups. Panels (b) and (c) of Figure 2.4 show ELF and POL at the grid-cell level. Visually, it is clear that the spatial variation in ELF and POL are quite different from the spatial variation in ethnic remoteness. In fact, the cell-level correlation between ethnic remoteness and ELF is only 0.09, and the corresponding correlation with POL is 0.13.

2.2.5 Ethnic Complementarity

One additional dimension of ethnicity that may matter for the relation between trade and conflict is ethnic complementarity. This concept aims to capture how much different ethnicities depend on each other and how likely they are to engage in productive cooperation. Stronger interethnic complementarities might lower the barriers to reaping the gains from trade, reducing the risk of conflict (Jha, 2013). On the other hand, the possibility to trade might disrupt and weaken the historic interdependence between ethnicities, increasing the risk of conflict. As measures of this interdependency, we use the concepts of ethnic specialization, kinship tightness, and segmentary lineage.

Ethnic specialization. Ethnic specialization measures the extent to which occupational specialization traditionally ran along ethnic lines. The idea is that if different ethnic groups specialize in different activities, they depend more on each other and they are more complementary to each other. To get a measure of ethnic specialization at the cell level, we combine information of the traditional occupational activity by ethnicity with the ethnic composition of grid cells. Denote by x_n^q the share of ethnic group n traditionally employed in occupation q, where the data on occupational activity come from the Ethnographic Atlas (Murdock, 1967b). Combining this with the ethnic composition of each grid-cell, we can determine the share of cell ℓ traditionally employed in occupation q, $x_\ell^q = \sum_n s_{n\ell} x_n^q$.[16] Following Krugman (1991), we can define the specialization of ethnic group n as $\sum_q \mid x_n^q - x^q \mid$, where x^q is the share of the country's population traditionally employed in

[16] As mentioned before, we use ethnicities and languages interchangeably. However, since occupational composition is measured by ethnicity, and cell composition is measured by language, we need an explicit mapping between ethnicities and languages. For that mapping, we rely on the work of Giuliano and Nunn (2018a).

occupation q. The extent of ethnic specialization of cell ℓ is then

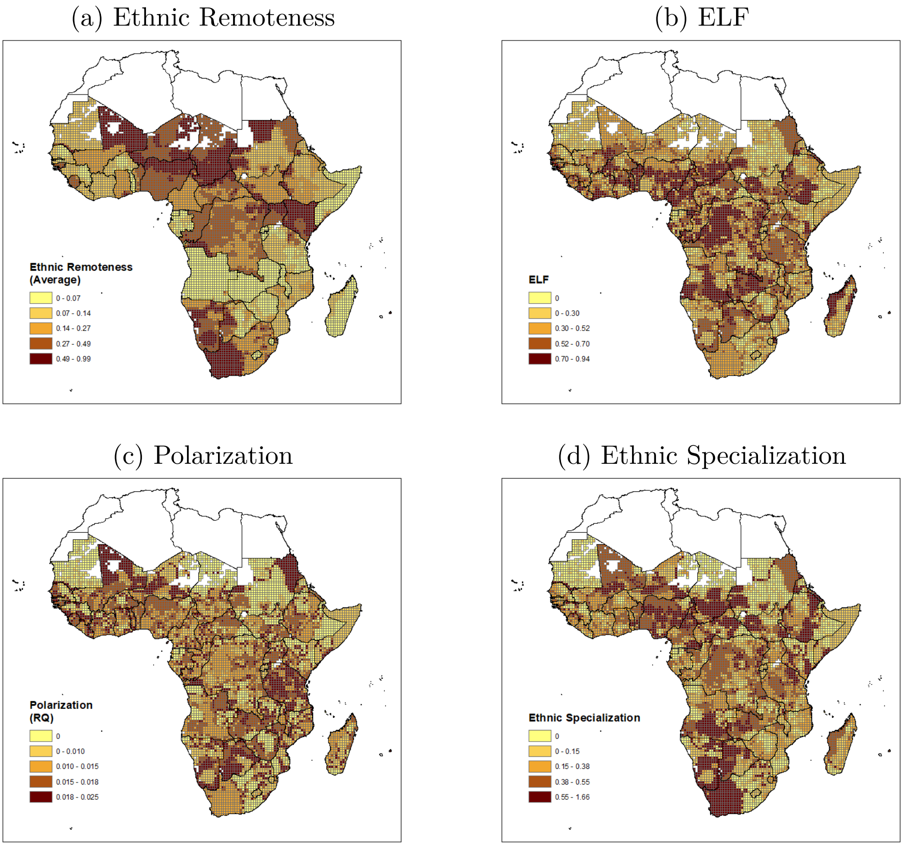
$$ES_\ell = \sum_n s_{\ell n} \sum_q \mid x_n^q - x^q \mid \tag{2.10}$$

The index is between 0 (no specialization along ethnic lines) and 2 (maximum specialization along ethnic lines). For ease of interpretation of the coefficients, we standardize ES_ℓ to have mean 0 and standard deviation 1. Panel (d) of Figure 2.4 shows a map of ethnic specialization at the local level. The correlation between ethnic remoteness and ethnic remoteness is 0.26.

Kinship tightness. As argued by Enke (2019), the looser the kinship links in a society, the easier it is to cooperate with distant strangers. In our view, ethnic groups are more complementary if they are able to reap the benefits from productive collaboration between them. Hence, the greater the kinship tightness of a cell, the lower the cell's ethnic complementarity. To measure a cell's kinship tightness, we use data on the kinship tightness by ethnicity from Enke (2019), and take the population-weighted average of the cell's different ethnic groups. Panel (a) of Appendix Figure 2.A.1 shows a cell-level map of kinship tightness. The correlation with ethnic remoteness is 0.11.

Segmentary lineage. Segmentary lineages are groups of people that trace their ancestry to a common founder. When an ethnic group is organized along segmentary lineages, it is less likely to form associations with other ethnicities, and it is more likely to engage in violent conflict (Moscona et al., 2020). As such, a cell populated by ethnicities that organize along segmentary lineages will experience a low level of ethnic complementarity. To measure a cell's segmentary lineage, we use ethnicity-level data on segmentary lineages from Moscona et al. (2020) and take its cell-level population-weighted average. Panel (b) of Appendix Figure 2.A.1 shows a map of segmentary lineage. The correlation with ethnic remoteness is -0.24.

Figure 2.4. Ethnic Remoteness, Ethnic Diversity, and Ethnic Specialization



Notes: Panel (a) plots ethnic remoteness, which measures the average ethnic distance between a random resident of the cell and a random resident of the country. Panel (b) plots the ethnolinguistic fractionalization index, which measures the probability that two randomly drawn individuals of a cell pertain to different ethnic groups. Panel (c) plots the ethnolinguistic polarization index. Panel (d) plots the ethnic specialization index, which measures the extent to which occupational specialization runs along ethnic lines.

2.2.6 Other Control Variables

As weather shocks are important predictors of conflict (Burke et al., 2015; Miguel et al., 2004; Ciccone, 2011), we control for both temperature and rainfall shocks. Following recent work, we use standardized deviations in rainfall and temperature (Hidalgo et al., 2010; Armand et al., 2020). The rainfall data are drawn from the CHIRPS dataset (Funk et al., 2014), while the temperature data come from the ERA reanalysis data (Muñoz-Sabater et al., 2021). The disease environment is also a predictor of conflict (Cervellati et al., 2022). Data on malaria suitability are drawn from Kiszewski et al. (2004), made available in raster format by McCord and Anttila-Hughes (2017). Data on crop suitability and Tse Tse fly suitability come from the FAO.

2.3 Ethnic Remoteness, Trade Access, and Conflict

Our primary objective is to explore the role of ethnic remoteness in mediating the relation between trade access and conflict. Ethnically more remote locations may face hurdles to fully participate in trading networks, possibly generating a relative increase in conflict in the wake of a trade agreement that improves access to foreign markets. In addition to ethnic remoteness, there may also be a role for ethnic diversity and ethnic complementarity. Ethnically more diverse locations may find it harder to share the benefits from a positive trade shock, leading to a relatively greater risk of conflict. Ethnically more complementary locations may witness either more conflict (if improved trade access weakens ethnic interdependence) or less conflict (if ethnic interdependence facilitates collaboration in the wake of improved trade access).

2.3.1 Cell-Level Regression Specification

Our main specification regresses cell-level conflict severity in time t on the cell's degree of trade openness at time t and on the interaction of that trade openness with different measures related to the cell's ethnic makeup, controlling for cell and country-time fixed effects as well as for time-varying cell characteristics that may affect conflict. More specifically,

$$\log(y_{ict}+1) = \alpha AGOAccess_{ict} + AGOAccess_{ict}\mathbf{E}'_{ic}\beta + \mathbf{X}'_{ict}\gamma + \delta_{ic} + \eta_{ct} + u_{ict} \tag{2.11}$$

where y_{ict} is the number of fatalities in cell i of country c in year t, $AGOAccess_{ict}$ is the degree of trade openness of cell i in country c in year t as defined in (2.1), $\mathbf{E}_{ic}$ is a vector of time-invariant cell-level variables related to ethnicity (ethnic remoteness, ethnic diversity, ethnic complementarity) which we interact with the cell's degree of trade openness in year t, $\mathbf{X}_{ict}$ is a vector of cell-level time-varying characteristics (weather shocks), δ_{ic} are cell fixed effects, η_{ct} are country-time fixed effects, and u_{ict} is an idiosyncratic error term. By using cell and country-time fixed effects, we address several concerns. Cell fixed effects absorb all time-invariant cell characteristics that might affect conflict. Country-time fixed effects absorb all characteristics that vary across countries and time, such as time-varying country characteristics that determine selection into the AGOA program. We always correct standard errors for spatial correlation within a 500 km radius and for infinite serial correlation following Conley (1999) and Hsiang (2010).[17]

Both of our treatment variables, $AGOAccess_{ict}$ and $AGOAccess_{ict}\mathbf{E}'_{ic}$, are continuous in nature. Specifically focusing on the second treatment variable, it consists of $AGOA_{ct} \times Proximity_{ic} \times \max_{j \in J}\{Production_{icj}\} \times \mathbf{E}'_{ic}$, where $Proximity_{ic}$ and $\mathbf{E}'_{ic}$ are continuous. The absence of binary treatment variables precludes using a standard dynamic differences-in-differences event study approach. This explains our choice of empirical specification (3.1).

2.3.2 Ethnic Remoteness Weakens the Peace Dividend from Trade

Ethnic diversity, ethnic remoteness, and ethnic complementarity. Table 2.1 reports results from estimating equation (3.1) using conflict data from UCDP. Column (1) shows that a higher degree of trade openness is associated with lower levels of conflict. Column (2) adds an interaction of trade openness with ethnic remoteness, measured as the linguistic distance between a random individual of the cell and a random individual of the country. As can be seen, ethnic remoteness diminishes the benign effect of trade openness on conflict. That is, ethnically remote cells reap a smaller peace dividend from trade openness.

Columns (3) and (4) add interaction terms between trade openness and the cell's ethnic diversity, measured as either ethnic fractionalization or ethnic polarization. These additional interactions are statistically insignificant. Columns (5) through (7) add interaction terms between trade openness and

[17] The correction of SEs for spatial and temporal correction is implemented using code from Fetzer (2020). The recent literature has usually allowed a spatial correlation of SEs within the distance of 100 km (see e.g. Armand et al., 2020) to 500 km (see e.g. Berman et al., 2017, and McGuirk and Burke, 2020). We choose the more demanding 500 km cutoff.

different measures of ethnic complementarity. Here as well, none of these additional interaction terms are statistically significant. Columns (3) to (7) do not affect our main coefficient of interest: the interaction of trade openness with ethnic remoteness continues to yield a positive and statistically significant coefficient at the 1% level, with a magnitude that is stable. The magnitude of the impact of ethnic remoteness on conflict is meaningful. Taking column (2) as our preferred specification, a one standard deviation increase in ethnic remoteness in a cell that is fully open to trade increases the fatalities from conflict by 5.3%. The corresponding number when going from the ethnically least remote cell to the ethnically most remote cell is a predicted increase in fatalities by 20.0%.

Table 2.1. AGOA and Conflict: Ethnic Remoteness

	Intensity of Conflict from UCDP						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
AGOAccess	-0.046***	-0.108***	-0.102***	-0.097***	-0.106***	-0.115***	-0.108***
	(0.009)	(0.017)	(0.017)	(0.017)	(0.016)	(0.030)	(0.020)
AGOAccess × ER		0.202***	0.206***	0.210***	0.206***	0.201***	0.202***
		(0.046)	(0.048)	(0.049)	(0.053)	(0.045)	(0.047)
AGOAccess × ELF			-0.017				
			(0.022)				
AGOAccess × POL				-0.108			
				(0.080)			
AGOAccess × Specialization					-0.013		
					(0.042)		
AGOAccess × Kinship						0.017	
						(0.054)	
AGOAccess × Segmented							-0.000
							(0.014)
Observations	269497	269497	269497	269497	269497	269497	269497

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (fatalities + 1), where fatalities is based on data from UCDP. The unit of observation is the PRIO GRID cell (resolution 0.5 × 0.5 decimal degrees, approximately 55km × 55km at the equator). All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. The sample includes 9,293 grid-cells spread across 48 sub-Saharan African countries for the period of 1989–2017. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Alternative measures of exposure to AGOA. Table 2.2 shows robustness to alternative measures of exposure to AGOA. First, columns (1) and (2) report results for a broader definition of AGOA as defined in equation (2.2). This definition measures a cell's exposure as proximity to the closest major port, without taking into account whether the cell produces any

AGOA-eligible products. Next, columns (3), (4), (5) and (6) report results for a narrower definition of AGOA as defined in equation (2.3). In addition to requiring a cell to produce an AGOA-eligible product, it makes exposure conditional on the country exporting that good to either the world or the U.S. in the pre-AGOA period. Finally, columns (7) and (8) follow equation (2.4) by defining exposure based on whether a cell has the adequate suitability conditions to produce an AGOA-eligible product.[18] As can be seen, when using any of these alternative measures of exposure to AGOA, our results remain unchanged.

Table 2.2. AGOA and Conflict: Alternative Definitions of AGOA Exposure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AGOAGeo	-0.144***	-0.176***						
	(0.031)	(0.033)						
AGOAGeo × ER		0.145***						
		(0.039)						
AGOAExp (World)			-0.049***	-0.115***				
			(0.009)	(0.018)				
AGOAExp (World) × ER				0.214***				
				(0.048)				
AGOAExp (US)					-0.054***	-0.114***		
					(0.011)	(0.020)		
AGOAExp (US) × ER						0.233***		
						(0.066)		
AGOASuit							-0.031***	-0.073***
							(0.009)	(0.017)
AGOASuit × ER								0.131***
								(0.045)
Observations	269497	269497	269497	269497	269497	269497	269497	269497

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (fatalities + 1), where fatalities is based on data from UCDP. Columns (1) and (2) use the broad definition of AGOA exposure without requiring the production of AGOA-eligible goods as defined in equation (2.2). Columns (3), (4), (5), and (6) use a narrow definition of AGOA that takes into account if the country has export capacity in AGOA-eligible goods to either the rest of the world or the U.S. as defined in equation (2.3). Columns (7) and (8) measure AGOA exposure conditional on a location's land being suitable for AGOA-eligible crops as defined in equation (2.4). The unit of observation is the PRIO GRID cell (resolution 0.5 × 0.5 decimal degrees, approximately 55km × 55km at the equator). All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. The sample includes 9,293 grid-cells spread across 48 sub-Saharan African countries for the period of 1989–2017. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Different types of goods. In Appendix Section 2.B.2, we investigate whether there are differences in the effect of AGOA exposure between cells that produce different types of AGOA-eligible goods. We distinguish between four types of goods: crops, textiles and apparel, other manufactured goods, and oil and minerals. For all four types, there is a peace dividend from improved trade access. However, only for crops do we find that ethnically

[18] Recall that for crops, suitability is defined as land suitability according to FAO data, whereas for other goods, suitability is defined as actually producing the goods.

remote cells experience a smaller drop in conflict. The absence of statistically significant results for non-agricultural products might be due to the small number of cells in that category. As can be seen in Appendix Tables B5 and B6, almost all treated cells produce crops, and almost all cells that produce non-agricultural AGOA-eligible goods also produce AGOA-eligible crops.

ACLED conflict data. As an alternative to the UCDP conflict data, we rerun the same regressions using conflict data based on ACLED in Appendix Section 2.B.2. As a reminder, the ACLED dataset is based on a broader definition of conflict, as it includes civil and communal conflicts, violence against civilians, and rioting and protesting. However, it only includes three years of pre-AGOA data. Table B15 uses the exact same specifications as Table 2.1, with the exception of the dependent variable. Our main result is unchanged: openness reduces conflict, but this benign effect is smaller in cells that are more ethnically remote from the rest of the country.

Robustness to environmental variables. Some variables may affect both a cell's ethnic remoteness and the degree of conflict it suffers. Because we include cell fixed effects, this is only an issue if these factors affect not just the level of conflict but also the change in conflict following accession to AGOA. One example would be if ethnically remote groups reside on marginal land, forcing them to rely on subsistence activity that does not lend itself to taking advantage of trade openness. Consistent with this, column (2) of Table 2.3 shows that cells that are unsuitable for crops benefit from a smaller peace dividend from AGOA. However, our main result does not change: the effect of ethnic remoteness, interacted with trade openness, is still positive, statistically significant at the 1% level, and of a similar magnitude.

Another example would be if areas with high incidence of malaria and other infectious diseases have more remote ethnic groups, because the disease environment incentivizes groups to isolate themselves. If a higher disease incidence also limits the gains from trade, then we should control for the interaction of the disease environment with AGOA.[19] Columns (3) and (4) of Table 2.3 report results when controlling for interactions with malaria and tsetse fly suitability. As expected, cells with higher malaria incidence get a smaller reduction in conflict after the AGOA trade shock. In contrast,

[19]See Cervellati et al. (2022) for evidence on the effect of malaria suitability on conflict in Africa.

cells with higher tsetse fly suitability show no difference. Again, our main finding is unchanged: ethnic remoteness weakens the peace dividend from trade liberalization.

Table 2.3. AGOA and Conflict: Robustness to Environmental Variables

	Intensity of Conflict from UCDP			
	(1)	(2)	(3)	(4)
AGOAccess	-0.108***	-0.183***	-0.114***	-0.099***
	(0.017)	(0.027)	(0.018)	(0.017)
AGOAccess × ER	0.202***	0.185***	0.193***	0.195***
	(0.046)	(0.047)	(0.047)	(0.046)
AGOAccess × Crop Unsuitability		0.016***		
		(0.004)		
AGOAccess × Malaria Suitability			0.022***	
			(0.008)	
AGOAccess × Tsetse Suitability				-0.008
				(0.006)
Observations	269497	269497	269497	269497

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (fatalities + 1), where fatalities is based on data from UCDP. The unit of observation is the PRIO GRID cell (resolution 0.5 × 0.5 decimal degrees, approximately 55km × 55km at the equator). All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. The sample includes 9,293 grid-cells spread across 48 sub-Saharan African countries for the period of 1989–2017. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Robustness to different measures of ethnic diversity. When exploring the interaction between a cell's openness and its ethnic diversity in Table 2.1, we relied on standard measures of fractionalization and polarization. Table B7 considers a number of alternative measures of diversity.[20]

First, in column (1) we use the Greenberg index, a generalization of the fractionalization index that takes into account the linguistic distances between the different ethnic groups (Greenberg, 1956; Desmet et al., 2009a):

$$GI_\ell = \sum_n \sum_m s_{\ell n} s_{\ell m} d_{nm}. \tag{2.12}$$

[20] Appendix Figure 2.A.2 shows maps of these alternative indices.

This index measures the average linguistic distance between two randomly drawn individuals of cell ℓ. Second, in columns (2) and (3) we use the standard fractionalization index, but now define languages at different levels of coarseness. Take the example of Chad: at the finest level, the country has 135 ethnic groups, corresponding to its 135 languages, whereas at the coarsest level, there are two ethnic groups, corresponding to the Nilo-Saharan and the Afro-Asiatic language family. Generalizing this example, Desmet et al. (2012) define ethnic groups at 15 different levels of coarseness, yielding 15 corresponding fractionalization indices, $ELF_\ell^{15}, \ldots, ELF_\ell^{1}$. Columns (2) and (3) use ELF_ℓ^{2} (more coarse) and ELF_ℓ^{9} (less coarse). Third, in column (4) we use a generalization of the polarization index that takes into account linguistic distance between the different groups, following Esteban and Ray (1994):

$$POL_\ell^{er} = \sum_n \sum_m s_{\ell n}^2 s_{\ell m} d_{nm}. \tag{2.13}$$

The interaction of these alternative measures of diversity with AGOA yield negative coefficients, indicating that cells that are more diverse benefit from a larger peace dividend. However, these results are not robust to using the ACLED conflict data (Table B17). More importantly, in both Tables B7 and B17 the main coefficient of interest on the interaction between AGOA openness and ethnic remoteness continues to be negative and statistically highly significant. The weaker peace dividend from AGOA in ethnically remote cells is a robust finding.

Generalized System of Preferences. Before accession to AGOA in 2001, the least developed countries (LDCs) in our sample already benefited from improved access to the U.S. market through the expansion of the Generalized System of Preferences (GSP) in 1997. Of the 48 countries in sub-Saharan Africa, 28 are in this group. It is important to explore whether differentiating between the impact of both shocks affects our results.

By analogy with exposure to AGOA, we define a cell's exposure to GSP as the product of three factors: a binary variable that indicates whether the cell's country is part of GSP, a continuous variable that measures proximity to the nearest major port, and a binary variable that indicates whether the cell produces GSP-eligible goods. Introducing a separate role for GSP also requires redefining exposure to AGOA in the LDCs. We say that a cell in an LDC receives an additional shock from AGOA only if it produces goods that are on the AGOA-eligible list but not on the GSP-eligible list.

To see the effect of introducing a separate channel for GSP, we run our baseline specification, but control for GSP exposure both as a separate term and interacted with ethnic remoteness. Table B13 shows that GSP did not have a statistically significant impact on conflict, though the signs on the coefficients are consistent with a peace dividend that is smaller in ethnically remote locations. The absence of a statistically significant effect may reflect the limited export capacity of LDCs to the U.S. (Figure 2.3, panel (c)). More importantly, controlling for GSP as a separate shock does not change our main findings. AGOA continues to have a benign effect of conflict, but less so in ethnically remote locations.

Robustness to specialization. Another concern is that ethnic remoteness might correlate positively with specialization in non-tradable or import-competing sectors. If so, this would limit, or even overturn, the gains from trade, and hence the peace dividend. For want of cell-level data on sectoral composition we cannot run this robustness check here. However, in Section 4.2, where we show results from individual-level regressions of income shocks on ethnic remoteness, we are able to control for an individual's profession. As we will see, doing so does not affect our key finding.

Ethnic remoteness from the dominant group. Rather than considering ethnic remoteness from the rest of the country, we consider ethnic remoteness from the country's dominant group for the same baseline specifications of Table 2.1. The results, reported in Table 2.4, confirm our previous conclusions. Whether we measure ethnic remoteness as distance to the country or to the dominant group, it lowers the peace dividend from trade openness.

Robustness to alternative transformations of the dependent variable. In order not to lose locations with no conflict, in our baseline analysis we use $\log(y_{ict} + 1)$ as the dependent variable, where y_{ict} is the number of fatalities in cell i of country c in year t. In Appendix Table B8 we explore alternative ways to transform the conflict data. One such alternative is to use the inverse hyperbolic sine transformation, $\log(y + \sqrt{y^2 + 1})$ and another is to use $\log(y_{ict} + 0.5)$. As can be seen, our findings do not change. We could also ignore the intensive margin by defining conflict as a binary variable that takes the value of 1 if the number of fatalities is greater than 0. Doing so does not change the results.

Table 2.4. AGOA and Conflict: Ethnic Remoteness from Dominant Group

	Intensity of Conflict from UCDP						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
AGOAccess	-0.046***	-0.080***	-0.075***	-0.069***	-0.080***	-0.083***	-0.078***
	(0.009)	(0.012)	(0.013)	(0.013)	(0.012)	(0.025)	(0.014)
AGOAccess × ER^{dom}		0.123***	0.125***	0.127***	0.123***	0.123***	0.123***
		(0.029)	(0.030)	(0.031)	(0.032)	(0.029)	(0.029)
AGOAccess × ELF			-0.014				
			(0.022)				
AGOAccess × POL				-0.100			
				(0.080)			
AGOAccess × Specialization					0.002		
					(0.039)		
AGOAccess × Kinship						0.007	
						(0.055)	
AGOAccess × Segmented							-0.003
							(0.014)
Observations	269497	269497	269497	269497	269497	269497	269497

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (fatalities + 1), where fatalities is based on data from UCDP. The unit of observation is the PRIO GRID cell (resolution 0.5 × 0.5 decimal degrees, approximately 55km × 55km at the equator). All specifications include a constant, and controls for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. The sample includes 9,293 grid-cells spread across 48 sub-Saharan African countries for the period of 1989–2017. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.4 Ethnic Remoteness, Trade, and Income

Our findings so far are consistent with an opportunity cost view of conflict. Indeed, if AGOA leads to gains from trade, then the ensuing higher income increases the opportunity cost of engaging in conflict. In addition, if ethnic remoteness acts as a barrier to reaping the full income benefits from trade liberalization, then the peace dividend should be weaker in ethnically more remote locations.

This opportunity cost interpretation assumes that the AGOA trade shock increases income, but less so in ethnically more remote locations. However, so far, we have not shown any results based on income. To test the consistency of the income channel with the data, we start by using the exact same cell-level regression specification as before, with the difference that we look at the effect of AGOA on income (as proxied by luminosity), rather than on conflict. We then use individual-level data from different waves of the Afrobarometer to see whether the ethnic barrier interpretation also holds at the individual level. We explore whether ethnically more remote individuals suffer negative income shocks when exposed to trade, compared to individuals that are ethnically less distant.

2.4.1 Ethnic Remoteness Weakens the Income Gains from Trade

In this section we examine the effects of AGOA and its interaction with ethnic remoteness on income, as proxied by luminosity. While sub-national statistical data on income are scarce, especially in the context of developing countries, a large number of papers pioneered by Henderson et al. (2012) have shown nightlight measured by satellites to provide a good proxy income.[21]

Table 2.5. AGOA and Luminosity: Ethnic Remoteness

	Income Proxied by Luminosity						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
AGOAccess	0.366***	0.464***	0.475***	0.485***	0.476***	0.387***	0.452***
	(0.028)	(0.041)	(0.047)	(0.047)	(0.043)	(0.066)	(0.048)
AGOAccess × ER		-0.321***	-0.315***	-0.307***	-0.299***	-0.337***	-0.317***
		(0.085)	(0.085)	(0.085)	(0.087)	(0.086)	(0.085)
AGOAccess × ELF			-0.031				
			(0.055)				
AGOAccess × POL				-0.209			
				(0.213)			
AGOAccess × Specialization					-0.090		
					(0.097)		
AGOAccess × Kinship						0.189	
						(0.130)	
AGOAccess × Segmented							0.021
							(0.042)
Observations	241072	241072	241072	241072	241072	241072	241072

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (nighttime light + 1). The unit of observation is the PRIO GRID cell (resolution 0.5 × 0.5 decimal degrees, approximately 55km × 55km at the equator). All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. The sample includes 8,670 grid-cells spread across 48 sub-Saharan African countries for the period of 1992–2017. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We take the same estimating equation (3.1) as before, but replace y_{ict} by luminosity. Table 2.5 reports our main results. We find what we expect: in all columns, the AGOA trade shock increases income, but less so in ethnically

[21]See Michalopoulos and Papaioannou (2018) for a review of the literature that has used luminosity data as a proxy for economic development.

Table 2.6. AGOA and Luminosity: Alternative Definitions of AGOA Exposure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AGOAGeo	1.518***	1.592***						
	(0.107)	(0.108)						
AGOAGeo × ER		-0.338***						
		(0.075)						
AGOAExp (World)			0.361***	0.466***				
			(0.029)	(0.042)				
AGOAExp (World) × ER				-0.341***				
				(0.086)				
AGOAExp (US)					0.249***	0.332***		
					(0.041)	(0.052)		
AGOAExp (US) × ER						-0.321***		
						(0.108)		
AGOASuit							0.214***	0.339***
							(0.034)	(0.044)
AGOASuit × ER								-0.387***
								(0.083)
Observations	241072	241072	241072	241072	241072	241072	241072	241072

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (nighttime light + 1). Columns (1) and (2) use the broad definition of AGOA exposure without requiring the production of AGOA-eligible goods as defined in equation (2.2). Columns (3), (4), (5), and (6) use a narrow definition of AGOA that takes into account if the country has export capacity in AGOA-eligible goods to either the rest of the world or the U.S. as defined in equation (2.3). Columns (7) and (8) measure AGOA exposure conditional on a location's land being suitable for AGOA-eligible crops as defined in equation (2.4). The unit of observation is the PRIO GRID cell (resolution 0.5 × 0.5 decimal degrees, approximately 55km × 55km at the equator). All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. The sample includes 9,293 grid-cells spread across 48 sub-Saharan African countries for the period of 1989–2017. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

remote locations. When looking at some of the other interaction terms, none of them are statistically significant.

Table 2.6 considers alternative definitions of exposure to AGOA. Columns (1) and (2) define exposure based on proximity to a major port, without taking into account product eligibility. Columns (3) through (6) make exposure conditional not just on product eligibility, but also on the country's export capacity of the product. Columns (7) and (8) define exposure in terms of suitability to produce AGOA-eligible products, rather than on actual production. Our findings are robust to these alternative ways of defining exposure.

Table 2.7 controls for the interaction of AGOA openness with different environmental variables. If cells that are ethnically remote have land that is unproductive, that may limit their capacity to reap the gains from trade. Consistent with this, column (2) shows that cells that are more unsuitable for crop production experience smaller income gains from AGOA openness. Cells that have a worse disease environment may also be in a disadvantaged position to benefit from trade. Though columns (3) and (4) show negative impacts of the incidence of either malaria or the tsetse fly, the effects are not statistically significant. None of these additional interaction terms affect the main finding: the income gains from trade are smaller in ethnically remote locations.

As further robustness checks, Appendix Table B9 includes alternative measures of fractionalization and polarization, Appendix Table B10 considers alternative transformations of our dependent variable, and Appendix Table B14 differentiates between the GSP and the AGOA shocks. These additional exercises have no qualitative impact on our main coefficients of interest.

Table 2.7. AGOA, Luminosity and Remoteness: Controlling for Environmental Variables

	Income Proxied by Nighlight			
	(1)	(2)	(3)	(4)
AGOAccess	0.464***	0.622***	0.471***	0.472***
	(0.041)	(0.076)	(0.042)	(0.044)
AGOAccess × ER	-0.321***	-0.285***	-0.311***	-0.327***
	(0.085)	(0.085)	(0.085)	(0.085)
AGOAccess × Crop Unsuitability		-0.033**		
		(0.013)		
AGOAccess × Malaria Suitability			-0.024	
			(0.023)	
AGOAccess × Tsetse Suitability				-0.007
				(0.017)
Observations	241072	241072	241072	241072

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (nighttime light + 1). The unit of observation is the PRIO GRID cell (resolution 0.5 × 0.5 decimal degrees, approximately 55km × 55km at the equator). All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. The sample includes 9,293 grid-cells spread across 48 sub-Saharan African countries for the period of 1992–2017. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.4.2 Individual-Level Evidence

If ethnic remoteness acts as a barrier to reaping the income gains from trade, we would expect to find evidence for this mechanism not just at the cell level, but also at the individual level. In this section, we use data from the Afrobarometer to explore how the effect of the AGOA trade access shock on income depends on an individual's ethnic remoteness.

Empirical specification. We regress measures of an individual's income on the trade openness of the cell where she resides and on the interaction of that openness with the individual's ethnic remoteness from either the rest of the country or from the country's dominant group. More specifically,

$$I_{jeict} = \alpha\, AGOAccess_{ict} + AGOAccess_{ict}\mathbf{E}'_{jec}\beta_1 + AGOAccess_{ict}\mathbf{E}'_{ic}\beta_2 + \mathbf{X}'_{ict}\gamma + \delta_{ic} + \eta_{ct} + \theta_e + u_{jeict} \tag{2.14}$$

where I_{jeict} is a measure of the income of individual j of ethnicity e residing in cell i of country c at time t, $AGOAccess_{ict}$ is the degree of trade openness of cell i in country c at time t, $\mathbf{E}'_{jec}$ is a vector of individual-level variables related to ethnicity which we interact with the cell's degree of trade openness at time t, $\mathbf{E}'_{ic}$ is a vector of cell-level variables related to ethnicity which we also interact with the degree of openness, $\mathbf{X}_{ict}$ is a vector of cell-level time-varying characteristics, δ_{ic} are cell fixed effects, η_{ct} are country-time fixed effects, θ_e are ethnicity fixed effects, and u_{jeict} is an idiosyncratic error term.

Individual data. We use individual-level data from the 12 countries that were included in all six rounds of the Afrobarometer surveys conducted between 1999–2015. This includes the first round that was conducted between 1999 and 2001, before the entry into AGOA for most countries.[22] As proxies for income, we use two measures: food poverty and income poverty. These measures correspond to the questions: "Over the past year, how often, if ever, have you or your family gone without: enough food to eat / cash income?". We recode the responses to these questions as binary variables, that take the value 1 if individuals answer "just once or twice", "several times", "many times" or "always", and the value 0 if individuals answer "never".

When estimating whether the income shock of trade has a differential effect on individuals that are ethnically remote, we need to know where the individual resides and which ethnicity she belongs to. An individual's location determines the size of the trade liberalization shock, and an individual's ethnicity determines her remoteness to either the country's average or the country's largest group. The Afrobarometer provides an individual's GPS

[22] Table A3 lists the countries for which we have individual-level survey responses prior to the entry to AGOA. Apart from Mali and Tanzania, which were surveyed in the same year as their entry into AGOA, all the other 10 countries were surveyed before entry into AGOA. This includes Zimbabwe, which was never part of AGOA.

location and her language (which, as before, we use as a proxy for ethnicity).[23]

Ethnically remote individuals and food poverty. In developing countries, food poverty is often a more reliable measure of economic well-being than income (Meyer and Sullivan, 2003). Table 2.8 reports results for regressions of individual-level food poverty on trade openness, using specification (2.14). All our individual-level regressions include ethnic group fixed effects, which among other things purge any possible effects of within-group genetic diversity (Arbatlı et al., 2020). Column (1) shows that individuals that are ethnically remote experience more food poverty in the wake of trade liberalization. Columns (2) and (3) suggest that ethnic remoteness of the individual, rather than ethnic remoteness of the location, drives the increased food poverty effect of trade liberalization. In terms of magnitudes, taking column (3) as our preferred specification, a one standard deviation increase in an individual's ethnic remoteness in a cell that is fully open to trade increases food poverty by 4.2 percent. Overall, this provides support to the hypothesis that an individual's ethnic remoteness makes it more difficult to take advantage of trade liberalization.

Profession and other individual controls. One concern is that ethnically remote individuals might work in professions that benefit less from trade liberalization. Another concern is that ethnically remote individuals might have other specific characteristics that affect their capacity to take advantage of a positive trade shock. In columns (4)–(6) of Table 2.8 we control for an individual's profession, age and gender, as well as for whether she resides in a rural location. The results are unchanged: individuals that either are ethnically remote are more likely to suffer from food poverty in the wake of a positive trade shock.[24]

---

[23] Table B3 provides the summary statistics of the individual-level data. Table B4 provides the correlation between individual- and cell-level measures of ethnic remoteness.

[24] We use the following professional categories "Agriculture / farming / fishing / forestry", "Artisan or skilled manual worker", "Clerical or secretarial", "Don't know", "Housewife / home-maker", "Missing", "Never had a job", "Other", "Professional", "Retail / Shop", "Security services", "Student", "Supervisor / Foreman / Senior Manager", "Trader / hawker / vendor", and "Unskilled manual worker." Waves 4 and 5 do not include information on occupational categories, at least for the 12 countries in our sample. Hence results in columns 4–6 of Table 2.8 are based on waves 1, 2, 3 and 6.

Table 2.8. AGOA and Food Poverty

	Individual Food Poverty					
	(1)	(2)	(3)	(4)	(5)	(6)
AGOAccess	-0.110**	-0.101**	-0.113**	-0.152**	-0.150**	-0.167**
	(0.052)	(0.051)	(0.055)	(0.069)	(0.069)	(0.072)
AGOAccess × Indiv ER	0.172***		0.167***	0.200***		0.179***
	(0.046)		(0.050)	(0.044)		(0.050)
AGOAccess × Cell ER		0.120*	0.016		0.183***	0.076
		(0.061)	(0.077)		(0.060)	(0.081)
Individual Controls	No	No	No	Yes	Yes	Yes
Observations	114176	114176	114176	72112	72112	72112

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is based on the answer to the question: “Over the past year, how often, if ever, have you or your family gone without: Enough food to eat?". It is coded as 0 (if answer is never) or 1 (if answer is sometimes/several times/frequently/many times/always). The unit of observation is the individual. The sample is based on six rounds of the Afrobarometer surveys conducted between 1999–2015 comprising approximately between 17k and 22k individual per round spread across 12 countries (see Appendix 2.A.3 for full list of countries). Cell-level ethnic remoteness is for the cell which the individual resides. All regressions control for rainfall and temperature shocks, country-specific cell FE, country-specific year FE and individual ethnolinguistic group FE. Columns (4)-(6) include additional individual controls for professions, age bracket, gender, and rural location. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Robustness. Table 2.9 replicates the above table but uses income poverty as an alternative measure of an individual’s well-being. Focusing on column (3), we see that individuals that are ethnically remote from the country average experience a smaller decrease in income poverty in the wake of trade liberalization. Controlling for individual characteristics, such as profession and age, does not change these findings (columns (4) to (6)). Focusing on column (6), a one standard deviation increase in an individual’s ethnic remoteness in a cell that is fully open to trade increases the chance of income poverty by 4.6 percent.

Table 2.9. AGOA and Income Poverty

	Individual Income Poverty					
	(1)	(2)	(3)	(4)	(5)	(6)
AGOAccess	-0.078	-0.111*	-0.126*	-0.096	-0.120	-0.141*
	(0.073)	(0.066)	(0.068)	(0.078)	(0.074)	(0.074)
AGOAccess × Indiv ER	0.194***		0.166***	0.209***		0.182***
	(0.057)		(0.060)	(0.053)		(0.056)
AGOAccess × Cell ER		0.319***	0.193*		0.303**	0.170
		(0.093)	(0.102)		(0.129)	(0.137)
Individual Controls	No	No	No	Yes	Yes	Yes
Observations	108463	108463	108463	66500	66500	66500

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is based on the answer to the question: "Over the past year, how often, if ever, have you or your family gone without: A cash income?". It is coded as 0 (if answer is never) or 1 (if answer is sometimes/several times/frequently/many times/always). The unit of observation is the individual. The sample is based on six rounds of the Afrobarometer surveys conducted between 1999–2015 comprising approximately between 13k and 22k individuals per round spread across 12 countries (see Appendix 2.A.3 for full list of countries). Indiv ER refers to the individual ethnic remoteness from her fellow citizens in the country. Cell ER refers to cell-level ethnic remoteness of the cell in which the individual resides. All regressions control for rainfall and temperature shocks, country-specific cell FE, country-specific year FE and individual ethnolinguistic group FE. Columns (4)-(6) include additional individual controls for professions, age bracket, gender, and rural location. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.10 explores robustness to using alternative definitions of exposure to AGOA. As before, these include making exposure conditional on exports to the U.S. or using an exposure measure based on land suitability, rather than on actual production of AGOA-eligible goods. Our findings are unchanged: belonging to a group that is ethnically remote from the country average increases the chance of experiencing food poverty when trade with the U.S. is liberalized.

Appendix Table B21 uses distance from the dominant group rather than distance from the average group. As before, greater individual's ethnic remoteness to the dominant group increases the probability of going without food. Appendix Table B22 introduces additional cell-level interactions of ethnic diversity and ethnic complementarity with trade openness. Appendix Tables B23 and B24 add cell-level interactions with alternative measures of diversity and environmental variables. Our main result is robust to introducing these different variables. Appendix Tables B25–B26 show that these results are robust to measuring remoteness as distance to the dominant group. From these different exercises, we conclude that it is more difficult for ethnically remote individuals to reap the gains from trade. This is consistent with an interpretation that ethnic distance acts as a barrier that limits the benefits from trade openness.

Table 2.10. AGOA and Food Poverty: Alternative Definitions of AGOA Exposure

	Geo		X World		X US		Suitability	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: No individual controls								
AGOA	-0.302	-0.142	-0.072*	-0.113**	-0.011	-0.033	-0.013	-0.048*
	(0.400)	(0.419)	(0.043)	(0.055)	(0.029)	(0.033)	(0.032)	(0.029)
AGOA × Indiv ER		0.217***		0.161***		0.099**		0.133***
		(0.052)		(0.050)		(0.046)		(0.044)
AGOA × Cell ER		-0.010		0.020		0.046		0.043
		(0.083)		(0.077)		(0.070)		(0.068)
Observations	114176	114176	114176	114176	114176	114176	114176	114176
Panel B: Individual Controls								
AGOA	-0.291	-0.093	-0.100	-0.167**	0.004	-0.027	-0.018	-0.066**
	(0.491)	(0.473)	(0.064)	(0.072)	(0.038)	(0.037)	(0.043)	(0.033)
AGOA × Indiv ER		0.231***		0.176***		0.145***		0.165***
		(0.051)		(0.050)		(0.049)		(0.048)
AGOA × Cell ER		0.050		0.077		0.089		0.087
		(0.085)		(0.080)		(0.075)		(0.074)
Observations	72112	72112	72112	72112	72112	72112	72112	72112

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is based on the answer to the question: "Over the past year, how often, if ever, have you or your family gone without: Enough food to eat?". It is coded as 0 (if answer is never) or 1 (if answer is sometimes/several times/frequently/many times/always). The unit of observation is the individual. The sample is based on six rounds of the Afrobarometer surveys conducted between 1999–2015 comprising approximately between 17k and 22k individual per round spread across 12 countries (see Appendix 2.A.3 for full list of countries). Indiv ER refers to the individual ethnic remoteness from her fellow citizens in the country. Cell ER refers to cell-level ethnic remoteness of the cell in which the individual resides. All regressions in both panels control for rainfall and temperature shocks, country-specific cell FE, country-specific year FE and individual ethnolinguistic group FE. Regressions in Panel B control for additional individual controls, which include FEs for professions, age bracket, gender, and rural location. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.5 Conclusion

This paper explored how ethnicity affects the relation between trade liberalization and conflict. Exploiting the staggered implementation of the Africa Growth and Opportunity Act (AGOA), we found that improved trade access generates a peace dividend, but less so in locations that are ethnically remote from the rest of the country. Our findings are consistent with an opportunity cost view of participating in conflict. As the gains from trade raise the standard of living, it becomes more costly to engage in conflict. For this mechanism to be a potential explanation of our main result, we would expect more remote locations to benefit less from a positive income shock in the wake of AGOA. We would also expect ethnically more remote individuals to face higher barriers to reap the income gains from trade. Using high-resolution luminosity data as well as individual-level poverty data from Afrobarometer, we found evidence in support of these predictions. Overall, we conclude that ethnic remoteness acts as a barrier to participating in the global economy. In addition to geographic remoteness and sectoral specialization, ethnic remoteness should be a key concern when analyzing which locations and groups might be left behind by globalization.

Appendix

2.A Data

2.A.1 Data Sources for Cell-Level Regressions

Variable (Source)	Description
Basemaps (GMI)	Basemaps come from the Seamless Digital Chart of the World (Version 10.0), which accompanies the World Geodatasets data from Global Mapping International. The maps were created by the authors using ArcGIS® software by Esri®.
Conflict intensity (ACLED, UCDP)	We measure conflict using fatalities in each cell for a specific year. Data are obtained from two event-based databases: The Uppsala Conflict Data Program (UCDP) (Sundberg and Melander, 2013) for 1989–2017 and the Armed Conflict Location & Event Data Project (ACLED) (Raleigh et al., 2010) for 1997–2017.
Travel time to nearest port (IFPRI)	Travel time to nearest major port in hours in the year 2000. Source: HarvestChoice (2015)
Linguistic composition of cells (Desmet et al., 2020)	Distribution of language groups at the resolution of 5 km × 5 km from Desmet et al. (2020). They construct the data combining three sources of information: data on the spatial distribution of population from Landscan (Source: `http://web.ornl.gov/sci/landscan/`), data on the linguistic composition of countries from Ethnologue Version 17 (Lewis et al., 2014), and maps on the geographic distribution of 6,905 distinct languages from the World Language Mapping System (Version 17) produced by Global Mapping International (Source: `https://worldgeodatasets.com/language/`). Using this information, they then use an iterative proportional fitting algorithm to construct a comprehensive 0.05×0.05 grid-cell level dataset on the ethnolinguistic composition of the population for the entire globe.
Poverty (Afrobarometer)	The sample is based on individual level data from six rounds of the Afrobarometer surveys conducted between 1999 and 2015 comprising approximately between 13k and 22k individuals per round spread across 12 countries (see Appendix 2.A.3 for full list of countries). Food poverty: Based on the answer to the question: "Over the past year, how often, if ever, have you or your family gone without: enough food to eat?". It is coded as 0 (if answer is never) or 1 (if answer is just once or twice / several times / many times / always). Income poverty: based on the answer to the question: "Over the past year, how often, if ever, have you or your family gone without: A cash income?". It is coded as 0 (if answer is never) or 1 (if answer is just once or twice / several times / many times / always).

(continued on next page)

Variable (Source)	Description
Nightlight (DMSP-OLS)	Nighttime light emission per square kilometer for 1992–2013 from the DMSP-OLS Nighttime Lights Time Series v.4 (National Oceanic and Atmospheric Administration, 2014) were downloaded from https://eogdata.mines.edu/products/dmsp. For 2014–2017 we use the extend series of VIRSS data generated by Nechaev et al. (2021), which make VIRSS and OLS data comparable. When there are two satellites for the same year, we take the average between both.
Precipitation (CHIRPS)	Average amount of daily precipitation (in mm) in the cell, based on daily precipitation data provided by the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) database (Funk et al., 2014). CHIRPS provides 0.05 $\times$ 0.05 resolution satellite imagery supplemented with in-situ monitoring station data. To ensure comparability of the measure across cells, we use double-standardized rainfall deviations (Hidalgo et al., 2010). We first account for seasonal patterns by standardizing monthly rain totals by cell and month for the period 1989–2020. For each cell, these indicators are then summed up by year and standardized over the same period.
Crop Suitability (FAO)	Crop suitability index (class) for low input level rain-fed cereals based on the average climate of baseline period 1961–1990. Source: FAO and IIASA (2012)
Tse Tse fly suitability (FAO)	These data come from http://www.fao.org/geonetwork/srv/en/main.home?uuid=f8a4e330-88fd-11da-a88f-000d939bc5d8. We use the median number of species, which lies between 0 and 10, in a grid cell as a measure of Tse Tse suitability.
Malaria suitability (Kiszewski et al., 2004)	Data on malaria suitability are drawn from Kiszewski et al. (2004), made available in raster format by McCord and Anttila-Hughes (2017).
Temperature (ERA)	Yearly mean temperature (in degrees Celsius) in the cell, based on monthly meteorological statistics from ERA Reanalysis dataset (Muñoz-Sabater et al., 2021). Data are available for the period 1948–2020. To ensure comparability of the measure across cells, we use standardized temperature deviations, by restricting the standardization to the year level.
AGOA Eligibile Crops (FAO, USITC, M3)	We combine the following 3 data sources: *Cell-level crop location* – We identify the cell-level location of 175 crops from the M3 crops data (Ramankutty et al., 2008) available at the 5 minute $\times$ 5 minute grid level for the year 2000 (average during the period 1997–2003). *US tariff line data* – Data on AGOA-eligible US tariffs at the eight-digit level come from the US International Trade Commission (USITC) for the year 2000 (https://dataweb.usitc.gov/tariff). *Crop description* – Description of the crops are from the FAO available at https://uses.plantnet-project.org/en/FAO,_product_nomenclature.

(continued on next page)

Variable (Source)	Description
	We combine the 175 identifiable crops from Ramankutty et al. (2008), with the USITC tariff data for the year 2000 using the FAO crop descriptions. From the 175 crops we keep 72 crops which appear as the main product of at least one AGOA-eligible tariff line. The 72 eligible crops are: alfalfa, almond, apple, apricot, artichoke, asparagus, avocado, bambara, barley, bean, blueberry, broadbean, cabbage, carrots, cauliflower, cerealnes, cherry, chicory, citrusnes, clover, cotton, cowpea, cucumberetc, currant, date, fig, grape, grapefruit, greenbean, greenbroadbean, greencorn, groundnut, hazelnut, hop, lemonlime, linseed, melonetc, millet, mushrooms, mustard, nutnes, oilseednes, olive, onion, orange, papaya, peachetc, pear, pineapple, plum, potato, quince, rapseed, raspberry, rice, rootnes, rye, ryefor, safflower, soybean, spinach, strawberry, stringbean, sugarbeet, sunflower, tangetc, tobacco, tomato, vegetablenes, walnut, watermelon, wheat. Appendix 2.C.1 provides further details on how we match the crops from FAO to US tariff lines.
AGOA Eligibile Minerals (SNL, USITC)	Data on mines come from S & P Global - SNL Metals and Mining (https://www.marketplace.spglobal.com/en/datasets/snl-metals-mining-(19)). The database provides the geo-location of 33 minerals. From these we select the minerals which constitute the main product of at least one AGOA-eligible tariff lines. The AGOA-eligible minerals include: bauxite (aluminum), iron, silver, zinc, cobalt, manganese (ferromanganese), niobium, tungsten, and vanadium. We use the geolocation of the mines to identify cells containing at least one mine of any AGOA-eligible mineral. Appendix 2.C.2 provides further details on how we match the minerals from S & P Global to US tariff lines.
AGOA Eligibile Oil (PETRODATA)	Data on oil are based on the PETRODATA dataset (Lujala et al., 2007) which contains information on oil and gas fields throughout the world. It covers 884 records for onshore and 378 for offshore reserves of natural gas and crude oil during the period 1946–2003. It includes a shapefile of polygons representing petroleum fields, which lets us identify cells overlapping oil fields.

(continued on next page)

Variable (Source)	Description
AGOA-Eligible Manufacturing (D & B)	To match AGOA-eligible manufactured goods to cells, we proceed in four steps. *Step 1: Match tariff lines to NAICS codes.* Our AGOA tariff lines correspond to the year 2000. However, the industry-level data are available at the 2022 North America Industry Classification System (NAICS) code level. We link the tariff lines from the year 2000 with NAICS at the 4-digit level from 2022 to determine which industries contain AGOA-eligible tariff lines for textile, apparel or other manufacturing or agro industries. To do so, we first match the entire set of tariff lines from the year 2000 with the Standard Industrial Classification (SIC) codes using the concordance provided by Pierce and Schott (2012). We then use the SIC code of each tariff line in 2000 to link them with all possible NAICS2022 codes at the 4-digit level, using data from the U.S. Bureau of Labor Statistics: https://www.bls.gov/ces/naics/#3.2.3. This provides us with a dataset at the NAICS2022 level, indicating how many tariff lines are present in each industry and the share that are AGOA-eligible (with separate categories for apparel and textiles). *Step 2: Identify locations of AGOA-eligible industries.* We identify the locations of all industries that have at least one tariff line categorized under AGOA. We take the full set of industries from the publicly available listings in the Business Directory of Dun & Bradstreet (D&B), which utilizes the four-digit NAICS classification. We then select the matched industries and collect all the locations of these industries for sub-Saharan African countries from Dun & Bradstreet (D&B). This process yields a comprehensive list of locations pertaining to 93 industries at the industry-country level, producing 39,089 observations. *Step 3: Assign latitude and longitude to each location.* The previous step gives us 7,651 unique locations to geolocalize. We start by assigning each location to a city. Since not all locations in Dun & Bradstreet correspond to a city, we proceed as follows. First, we do a fuzzy match of our locations with the names of cities in the comprehensive version of the World City Database (WCD) constructed by Simple Maps (2023). This yields 3,191 matches (of which 1,759 exact matches). For the remaining unmatched locations, we conduct an individual Google search to manually assign them to a city. If a location corresponds to a neighborhood or a district of a city, we retain that information to enhance precision in geocoding. If we are unable to match a location with a city after the Google search, we return to D&B to retrieve the exact address of one of the companies in these locations. Finally, we are left with only 115 unmatched locations. For all the matched locations, we obtain longitude and latitude from either the WCD or the "Awesome Table" add-on in Excel. To ensure accuracy when using the add-on, we exclude locations whose latitude and longitude coincide with the centroid of the country. With this quality control measure, we remove 354 locations. *Step 4: Match with grid cells.* We perform a spatial join between our geocoded locations and grid cells to assign each AGOA-eligible industry location to a cell. For locations that fall between two cells, both cells are assigned to the location, which occurred in 14 cases. Additionally, 56 observations were not assigned a cell through the spatial join. We manually reviewed these observations, discarding 23 locations identified as clear errors.

(continued on next page)

Variable (Source)	Description
GSP variables	To create the GSP variables we follow the same procedure as for the AGOA variables. However, instead of focusing on AGOA-eligible tariff lines, we focus on GSP-eligible tariff lines in year 1997.
AGOA Suitability (FAO)	We obtain crop suitability data from the FAO's Global Agro-Ecological Zones (FAO and IIASA, 2012). This dataset provides estimates of suitability for individual crops at a 5 arc-minutes resolution for historical, current, and future conditions. Specifically, we use the suitability index, which takes values from 0–10,000 depending on how suitable each cell is (variable name is "Suitability index range (0 – 10,000); all land in grid cell"). We download the data selecting the following options: rain-fed (water supply is "rainfed"), high input intensity (input level is "high"), and without CO2 fertilization (CO2 fertilization is "Without CO2 Fertilization"). The measure is calculated for the period 1971–2000. Following Nunn and Qian (2011), we define a cell as suitable if it is classified in the database as being either "very suitable", "suitable", or "moderately suitable". In other words, a cell is suitable if it has a value greater or equal than 4,000 in the suitability index. From the selection of crops available from GAEZ, we choose those which are present in at least one of the tariff lines eligible under AGOA as the main product of the tariff. These crops are the following: alfalfa, barley, phaseolus bean, cabbage, carrot, citrus, cotton, cowpea, groundnut, foxtail millet, pearl millet, olive, onion, white potato, rapeseed, dryland rice, wetland rice, rye, soybean, sugarbeet, sunflower, tobacco, tomato, and wheat.
Pre-AGOA Exports (CEPII)	The CEPII-BACI dataset gives us the 6-digit product identifier and the corresponding bilateral country-level trade from 1995–2021 (Gaulier and Zignago, 2010). These were downloaded from http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=37 on June 19, 2023. We use the version 202301 last updated on February 1st, 2023. The exact downloading option chosen was called: HS92 (1995–2021).

2.A.2 AGOA Membership

Table A1. Years of access to AGOA

Country	AGOA years	No. of years
Angola	2004 – 2017	14
Benin	2001 – 2017	17
Botswana	2001 – 2017	17
Burkina Faso	2005 – 2017	13
Burundi	2006 – 2015	10
Cameroon	2001 – 2017	17
Cape Verde	2001 – 2017	17
Central African Republic	2001 – 2003; 2017	4
Chad	2001 – 2017	17
Comoros	2008 – 2017	10
DRC	2003 – 2010	8
Congo (ROC)	2001 – 2017	17
Cote d'Ivoire	2002 – 2004; 2011 – 2017	10
Djibouti	2001 – 2017	17
Eritrea	2001 – 2003	3
Ethiopia	2001 – 2017	17
Gabon	2001 – 2017	17
Gambia	2003 – 2014	12
Ghana	2001 – 2017	17
Guinea	2001 – 2009; 2011 – 2017	16
Guinea-Bissau	2001 – 2012; 2015 – 2017	15
Kenya	2001 – 2017	17
Lesotho	2001 – 2017	17
Liberia	2007 – 2017	11
Madagascar	2001 – 2009; 2014 – 2017	13
Malawi	2001 – 2017	17
Mali	2001 – 2012; 2014 – 2017	16
Mauritania	2001 – 2005; 2007 – 2008; 2010 – 2017	15
Mauritius	2001 – 2017	17
Mozambique	2001 – 2017	17
Namibia	2001 – 2017	17
Niger	2001 – 2009; 2014 – 2017	16
Nigeria	2001 – 2017	17
Rwanda	2001 – 2017	17
Sao Tome & Principe	2001 – 2017	17
Senegal	2001 – 2017	17
Seychelles	2001 – 2016	16
Sierra Leone	2001 – 2017	17
South Africa	2001 – 2017	17
South Sudan	2013 – 2014	2
Swaziland	2001 – 2014	14
Tanzania	2001 – 2017	17
Togo	2008 – 2017	10
Uganda	2001 – 2017	17
Zambia	2001 – 2017	17

Notes: This table reports the years in which the different sub-Saharan African countries enjoyed access to free trade with the U.S. under AGOA. Data are based on Appendix A of Fernandes et al. (2023). Our data stop in the year 2017, though AGOA might have continued to subsequent years.

Table A2. Years of access to Apparel and Textiles

Country	Apparel years	Textile Years	No. of years	No. of years
Benin	2005 – 2017	2005 – 2017	13	13
Botswana	2001 – 2017	2002 – 2017	17	16
Burkina Faso	2006 – 2017	2006 – 2017	13	13
Cameroon	2002 – 2017	2002 – 2017	16	16
Cape Verde	2002 – 2017	2002 – 2017	16	16
Central African Republic	2001 – 2003; 2017	2001 – 2003; 2017	4	4
Chad	2006 – 2017	2006 – 2017	12	12
Cote d'Ivoire	2003 – 2004; 2013 – 2017	2003 – 2004; 2013 – 2017	7	7
Ethiopia	2001 – 2017	2001 – 2017	17	17
Gambia	2008 – 2014	2008 – 2014	7	7
Ghana	2002 – 2017	2002 – 2017	16	16
Kenya	2001 – 2017	2001 – 2017	17	17
Lesotho	2001 – 2017	2001 – 2017	17	17
Liberia	2011 – 2017	2011 – 2017	7	7
Madagascar	2001 – 2009	2001 – 2009	9	9
Malawi	2001 – 2017	2001 – 2017	17	17
Mali	2003 – 2012	2003 – 2012	10	10
Mauritius	2001 – 2017	2007 – 2017	17	13
Mozambique	2002 – 2017	2002 – 2017	16	16
Namibia	2002 – 2017	2002 – 2017	16	16
Niger	2004 – 2009; 2014 – 2017	2004 – 2009; 2014 – 2017	13	13
Nigeria	2004 – 2017	2004 – 2017	14	14
Rwanda	2004 – 2017	2004 – 2017	14	14
Senegal	2002 – 2017	2002 – 2017	16	16
Sierra Leone	2004 – 2017	2004 – 2017	14	14
South Africa	2001 – 2017	-	17	0
Swaziland	2001 – 2014	2001 – 2014	14	14
Tanzania	2002 – 2017	2002 – 2017	16	16
Uganda	2001 – 2017	2001 – 2017	17	17
Zambia	2001 – 2017	2001 – 2017	17	17

Notes: This table reports the years in which the different sub-Saharan African countries enjoyed access to free trade with the U.S. under Apparel and Textiles categories. Data are based on Appendix A of Fernandes et al. (2023). Some countries, such as Burundi, Guinea and Mauritania, had access to AGOA but not to the categories of Apparel and Textiles. Our data go till the year 2017.

2.A.3 Afrobarometer Data

We use the 12 countries that were included in all 6 Afrobarometer rounds, spanning the period 1999 to 2015. This includes the first round of the Afrobarometer surveys conducted between 1999 and 2001, which for the vast majority of countries was before the entry into AGOA in the year 2001. Table A3 provides the information on the countries for which we have individual-level survey responses prior to the entry to AGOA. Apart from Mali and Tanzania, which were surveyed in the same year as AGOA entry, all the other 10 countries were surveyed before entry into AGOA. This includes Zimbabwe, which was never part of AGOA.

Table A3. Afrobarometer Round 1 and year of entry to AGOA

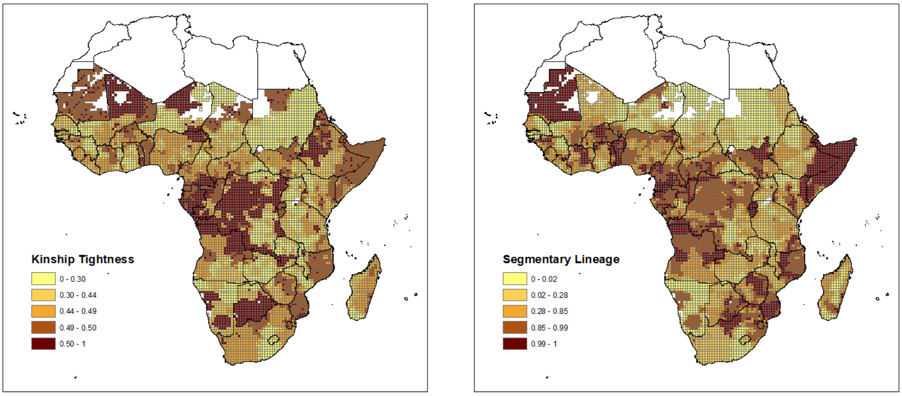
Country	AGOA entry	Survey Year
Botswana	2001	1999
Ghana	2001	1999
Lesotho	2001	2000
Malawi	2001	1999
Mali	2001	2001
Namibia	2001	1999
Nigeria	2001	2000
South Africa	2001	2000
Tanzania	2001	2001
Uganda	2001	2000
Zambia	2001	1999
Zimbabwe	NA	1999

Notes: This table provides the information on the countries for which we have individual-level survey responses prior to the entry to AGOA. Apart from Mali and Tanzania, which were surveyed in the same year as AGOA entry, all the other 10 countries were surveyed before entry into AGOA. This includes Zimbabwe, which was never part of AGOA.

2.A.4 Data Maps

Figure 2.A.1. Kinship Tightness and Segmentary Lineage

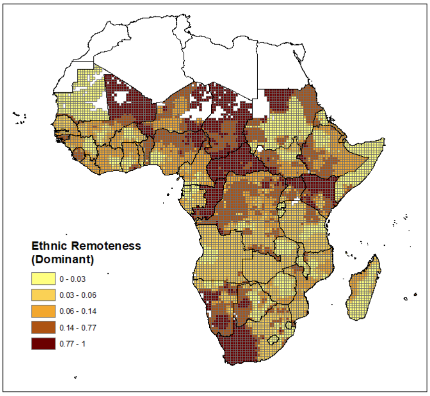
(a) Kinship Tightness **(b) Segmentary Lineage**



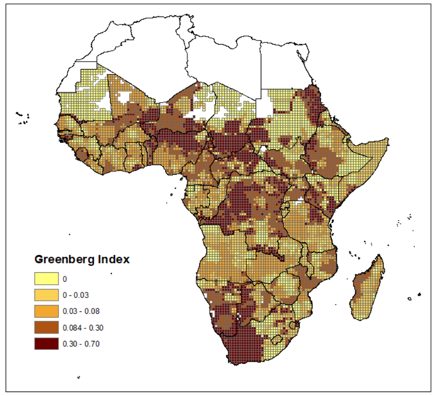
Notes: Panel (a) plots the average kinship tightness in a cell (Enke, 2019). Panel (b) plots the average segmentary lineage in a cell (Moscona et al., 2020). The distribution of ethnic groups is based on data from Desmet et al. (2020). See Appendix 2.A.1 for further details on data sources and variable definitions.

Figure 2.A.2. Alternative Ethnic Remoteness and Ethnic Diversity

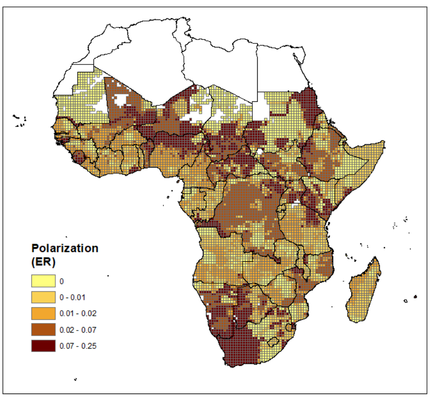
(a) Ethnic Remoteness from the Dominant Group



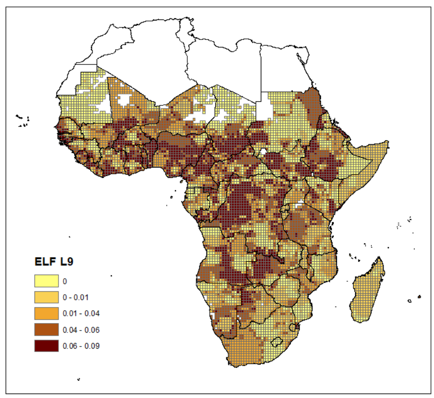
(b) Greenberg Index



(c) Polarization (ER)



(d) ELF Level 9



Notes: Panel (a) plots ethnic remoteness from the dominant group, which measures the average ethnic distance between a random resident of the cell and a random member of the most populous ethnic group in the country (equation 2.7). Panel (b) plots the Greenberg index, which measures the expected ethnic distance between any two random residents of the cell (equation 2.12). Panel (c) plots the Polarization index (equation 2.13), à la Esteban and Ray (1994). Panel (d) plots the fractionalization index at aggregation level 9, à la Desmet et al. (2012). The distribution of ethnic groups is based on data from Desmet et al. (2020). See Appendix 2.A.1 for further details on data sources and variable definitions.

2.B Additional Tables

2.B.1 Summary Statistics

Table B1. Summary Statistics (Cell-level)

Variable	Mean	Std. Dev.	Min.	Max.	N
Log(Fatalities + 1): UCDP	0.079	0.54	0	12.7	269497
Log(Fatalities + 1): ACLED	0.13	0.637	0	11.079	195153
Log(Luminosity + 1)	1.77	2.798	0	12.028	241072
Openness	0.863	0.102	0.095	1	269497
AGOAccess	0.278	0.418	0	1	269497
AGOAGeo	0.382	0.438	0	1	269497
AGOAExp (World)	0.271	0.415	0	1	269497
AGOAExp (US)	0.165	0.352	0	1	269497
AGOAccess (Crops)	0.268	0.414	0	1	269497
AGOAccess (Crops)/WExports	0.263	0.411	0	1	269497
AGOAccess (Crops)/USExports	0.147	0.336	0	0.998	269497
AGOAccess (Minerals/Oil)	0.037	0.181	0	0.998	269497
AGOAccess (Minerals/Oil)/WExports	0.028	0.158	0	0.998	269497
AGOAccess (Minerals/Oil)/USExports	0.02	0.134	0	0.995	269497
AGOAccess (Manufacturing)	0.031	0.168	0	1	269497
AGOAccess (Manufacturing)/WExports	0.031	0.168	0	1	269497
AGOAccess (Manufacturing)/USExports	0.027	0.159	0	1	269497
AGOAccess (Apparel)	0.02	0.137	0	0.998	269497
AGOAccess (Apparel)/WExports	0.02	0.137	0	0.998	269497
AGOAccess (Apparel)/USExports	0.02	0.136	0	0.998	269497
AGOASuit	0.275	0.415	0	1	269497
AGOASuit (Crops)	0.255	0.405	0	1	269497
ER	0.295	0.26	0	0.989	269497
$\mathrm{ER}^{\mathrm{dom}}$	0.274	0.358	0	1	269497
Ethnic Specialization	0.181	0.177	0	1	269497
Kinship Tightness	0.431	0.128	0	1	269497
Segmentary Lineage	0.533	0.417	0	1	269497
Greenberg	0.129	0.166	0	0.700	269497
ELF2	0.16	0.199	0	0.824	269497
ELF9	0.299	0.277	0	0.915	269497
ELF15	0.381	0.305	0	0.941	269497
POL	0.11	0.08	0	0.25	269497
$\mathrm{POL}^{\mathrm{er}}$	0.037	0.051	0	0.25	269497
Crop Unsuitability	5.411	1.548	1	9	269497
Malaria Suitability	0.28	0.98	-0.942	2.975	269497
TseTse Suitability	0.851	1.265	0	6	269497

2. Ethnic Remoteness Reduces the Peace Dividend from Trade Access

Table B2. Cell-Level Cross-Correlations

Variables	ER	ER Dom
ER	1.00	
ER Dom	0.85	1.00
Log (Fatalities + 1) UCDP	0.01	0.01
Log (Fatalities + 1) ACLED	0.02	0.01
Log (Luminosity + 1)	-0.01	-0.07
Openness	-0.26	-0.29
AGOAGeo	0.02	-0.00
AGOAccess	0.15	0.05
AGOAExp (World)	-0.11	-0.12
AGOAExp (US)	0.03	-0.04
AGOASuit	-0.18	-0.14
Ethnic Specialization	0.26	0.17
Kinship Tightness	0.11	0.13
Segmentary Lineage	-0.24	-0.20
Greenberg	0.49	0.36
ELF2	0.46	0.33
ELF9	0.22	0.15
ELF15	0.09	0.05
POL^{rq}	0.13	0.09
POL^{er}	0.54	0.41
Crop Unsuitability	0.24	0.22
Malaria Suitability	-0.01	-0.03
TseTse Suitability	-0.11	-0.06

Table B3. Individual-Level Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
		Panel A			
Food Poverty	0.51	0.5	0	1	114176
Openness	0.92	0.04	0.44	1	114176
AGOAOpen	0.74	0.37	0	1	114176
AGOAccess	0.54	0.5	0	1	114176
AGOAExp (World)	0.89	0.32	0	1	114176
AGOAExp (US)	0.70	0.46	0	1	114176
AGOASuit	0.81	0.39	0	1	114176
Cell ER	0.25	0.22	0.01	0.97	114176
Indiv ER	0.24	0.25	0	1	114176
Cell ER^{dom}	0.17	0.28	0	1	114176
Indiv ER^{dom}	0.16	0.34	0	1	114176
Female	0.5	0.5	0	1	114070
Rural	0.6	0.49	0	1	113846
Age	36.78	14.83	15	115	112896
		Panel B			
Income Poverty	0.76	0.42	0	1	108463
Openness	0.92	0.04	0.44	1	108463
AGOAOpen	0.78	0.34	0	1	108463
AGOAccess	0.51	0.5	0	1	108463
AGOAExp (World)	0.88	0.32	0	1	108463
AGOAExp (US)	0.68	0.47	0	1	108463
AGOASuit	0.81	0.39	0	1	108463
Cell ER	0.24	0.22	0.01	0.97	108463
Indiv ER	0.23	0.25	0	1	108463
Cell ER^{dom}	0.16	0.27	0	1	108463
Indiv ER^{dom}	0.16	0.34	0	1	108463
Rainfall Deviation	0.81	0.59	0	4.67	108463
Temperature Deviation	0.07	0.74	-2.13	2.54	108463
Female	0.5	0.5	0	1	108358
Rural	0.61	0.49	0	1	108121
Age	36.93	14.88	15	115	107194

Notes: Summary statistics for the individual-level data from six rounds of the Afrobarometer surveys. Panel A (Panel B) summarizes the sample for which the food poverty (income poverty) variable is available. These surveys were conducted between 1999–2015 comprising approximately between 17k and 22k individuals (Panel A) and 13k and 22k individuals (Panel B) per round spread across 12 countries (see Appendix 2.A.3 for full list of countries). The regressions in the paper use a gender dummy, which we display as a female dummy here. The regressions in the paper control for age categories rather than the age variable summarized here. See Section 3.4 and Appendix 2.A.1 for further details on data sources and variable definitions.

Table B4. Individual-Level Cross-Correlations

Variables	Indiv ER	Indiv ER^{dom}
Indiv ER^{dom}	0.89	-
Cell ER	0.80	0.60
Cell ER^{dom}	0.66	0.67

Notes: The sample includes 116,183 individual-level observations from six rounds of the Afrobarometer surveys conducted between 1999–2015 spread across 12 countries (see Appendix 2.A.3 for full list of countries). See Section 3.4 and Appendix 2.A.1 for further details on data sources and variable definitions.

Table B5. Share of Cells Producing Different Types of Goods, by Country

Country	Crops	Textiles	Apparel	Manufacturing	Minerals/Oil	Apparel & Textiles
Angola	0.75	0.00	0.00	0.01	0.06	0.00
Benin	1.00	0.00	0.02	0.04	0.04	0.02
Botswana	0.22	0.02	0.04	0.10	0.11	0.04
Burkina Faso	0.97	0.01	0.00	0.03	0.14	0.01
Burundi	0.89	0.00	0.00	0.11	0.17	0.00
Cameroon	0.69	0.01	0.01	0.03	0.13	0.01
Cape Verde	0.27	0.00	0.00	0.20	0.00	0.00
Central African Republic	0.52	0.00	0.00	0.00	0.00	0.00
Chad	0.51	0.00	0.00	0.00	0.06	0.00
Comoros	0.80	0.00	0.00	0.20	0.00	0.00
Congo (DRC)	0.57	0.00	0.00	0.01	0.05	0.00
Cote D'Ivoire	0.96	0.01	0.01	0.06	0.10	0.01
Djibouti	0.18	0.00	0.00	0.06	0.00	0.00
Eritrea	0.41	0.00	0.00	0.01	0.10	0.00
Ethiopia	0.70	0.01	0.01	0.01	0.02	0.01
Gabon	0.39	0.01	0.00	0.02	0.39	0.01
Ghana	0.94	0.03	0.01	0.10	0.13	0.03
Guinea	0.97	0.00	0.01	0.01	0.28	0.01
Guinea-Bissau	0.88	0.00	0.00	0.04	0.04	0.00
Kenya	0.87	0.02	0.02	0.05	0.03	0.03
Lesotho	0.95	0.14	0.05	0.14	0.00	0.14
Liberia	0.82	0.00	0.00	0.02	0.31	0.00
Madagascar	0.93	0.01	0.00	0.03	0.11	0.01
Malawi	0.92	0.02	0.00	0.10	0.19	0.02
Mali	0.46	0.00	0.00	0.00	0.03	0.00
Mauritania	0.19	0.00	0.00	0.00	0.05	0.00
Mauritius	0.33	0.33	0.33	0.33	0.00	0.33
Mozambique	0.88	0.01	0.00	0.03	0.06	0.01
Namibia	0.34	0.02	0.01	0.10	0.16	0.03
Niger	0.40	0.00	0.00	0.00	0.02	0.00
Nigeria	0.98	0.01	0.02	0.07	0.17	0.02
Republic of Congo	0.35	0.00	0.01	0.02	0.14	0.01
Rwanda	0.88	0.00	0.06	0.25	0.13	0.06
Senegal	0.92	0.01	0.01	0.05	0.10	0.01
Seychelles	0.00	0.00	0.00	0.11	0.00	0.00
Sierra Leone	0.87	0.00	0.00	0.05	0.34	0.00
South Africa	0.84	0.37	0.51	0.58	0.12	0.55
South Sudan	0.87	0.00	0.00	0.00	0.07	0.00
Swaziland	0.86	0.14	0.14	0.14	0.07	0.21
Tanzania	0.93	0.01	0.01	0.04	0.11	0.01
The Gambia	0.93	0.00	0.00	0.07	0.00	0.00
Togo	1.00	0.03	0.00	0.03	0.09	0.03
Uganda	0.95	0.03	0.02	0.04	0.13	0.03
Zambia	0.78	0.01	0.01	0.03	0.11	0.01
Average	0.67	0.03	0.04	0.06	0.09	0.04

Notes: This table presents the share of cells producing the different types of AGOA-eligible goods.

Table B6. Share of Cells Producing Both AGOA-Eligible Non-Crop Goods and AGOA-Eligible Crops, by Country

Country	Apparel & Textiles	Apparel	Textiles	Manufacturing	Minerals/Oil	All Non-Crop
Angola	.	.	.	0.86	0.81	0.82
Benin	1.00	1.00	.	1.00	1.00	1.00
Botswana	0.89	0.88	1.00	0.70	0.44	0.51
Burkina Faso	1.00	.	1.00	1.00	1.00	1.00
Burundi	.	.	.	1.00	1.00	1.00
Cameroon	1.00	1.00	1.00	1.00	0.54	0.57
Cape Verde	.	.	.	0.67	.	0.67
Central African Republic	.	.	.	1.00	1.00	1.00
Chad	.	.	.	1.00	0.81	0.81
Comoros	.	.	.	1.00	.	1.00
Congo (DRC)	1.00	1.00	1.00	1.00	0.76	0.78
Cote D'Ivoire	1.00	1.00	1.00	1.00	1.00	1.00
Djibouti	.	.	.	1.00	.	1.00
Eritrea	.	.	.	1.00	0.71	0.71
Ethiopia	1.00	1.00	1.00	1.00	1.00	1.00
Gabon	1.00	.	1.00	1.00	0.45	0.47
Ghana	1.00	1.00	1.00	1.00	1.00	1.00
Guinea	1.00	1.00	.	1.00	1.00	1.00
Guinea-Bissau	.	.	.	1.00	1.00	1.00
Kenya	1.00	1.00	1.00	1.00	1.00	1.00
Lesotho	1.00	1.00	1.00	1.00	.	1.00
Liberia	.	.	.	1.00	0.94	0.94
Madagascar	1.00	1.00	1.00	0.86	0.96	0.94
Malawi	1.00	.	1.00	1.00	1.00	1.00
Mali	1.00	1.00	1.00	1.00	0.86	0.87
Mauritania	.	.	.	0.00	0.07	0.07
Mauritius	1.00	1.00	1.00	1.00	.	1.00
Mozambique	1.00	1.00	1.00	1.00	1.00	1.00
Namibia	0.75	1.00	0.71	0.71	0.61	0.60
Niger	1.00	.	1.00	1.00	0.14	0.14
Nigeria	1.00	1.00	1.00	1.00	0.98	0.99
Republic of Congo	0.00	0.00	.	0.33	0.29	0.30
Rwanda	1.00	1.00	.	1.00	1.00	1.00
Senegal	1.00	1.00	1.00	1.00	0.70	0.77
Seychelles	.	.	.	0.00	.	0.00
Sierra Leone	.	.	.	1.00	0.92	0.93
South Africa	0.98	0.98	0.98	0.97	0.83	0.94
South Sudan	.	.	.	1.00	0.75	0.76
Swaziland	1.00	1.00	1.00	1.00	1.00	1.00
Tanzania	1.00	1.00	1.00	0.92	1.00	0.98
The Gambia	.	.	.	1.00	.	1.00
Togo	1.00	.	1.00	1.00	1.00	1.00
Uganda	1.00	1.00	1.00	1.00	1.00	1.00
Zambia	1.00	1.00	1.00	0.89	0.82	0.83
Average	0.97	0.98	0.98	0.94	0.79	0.84

Notes: This table shows the proportion of cells that produce AGOA-eligible non-crop goods and also produce AGOA-eligible crops. For instance, the value 0.89 in the 'Apparel & Textiles' column for Botswana indicates that 89% of cells producing textiles and apparel also produce crops. Missing values signify that no cells in the country produce goods in this category.

2.B.2 Robustness: Cell-Level Regressions

Baseline Definition of AGOA

Table B7. AGOA and Conflict: Different Diversity Measures

	(1)	(2)	(3)	(4)
AGOAccess	-0.108***	-0.105***	-0.101***	-0.116***
	(0.017)	(0.017)	(0.016)	(0.018)
AGOAccess × ER	0.280***	0.244***	0.213***	0.331***
	(0.066)	(0.058)	(0.051)	(0.072)
AGOAccess × Greenberg	-0.152**			
	(0.061)			
AGOAccess × ELF2		-0.082*		
		(0.043)		
AGOAccess × ELF9			-0.030	
			(0.027)	
AGOAccess × POL^{er}				-0.722***
				(0.205)
Observations	269497	269497	269497	269497

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (fatalities + 1), where fatalities is based on data from UCDP. The unit of observation is the PRIO GRID cell. All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. The sample includes 9,293 grid-cells spread across 48 sub-Saharan African countries for the period of 1989–2017. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2. Ethnic Remoteness Reduces the Peace Dividend from Trade Access

Table B8. AGOA and Conflict: Alternative Transformations of Dependent Variable

	Intensity of Conflict from UCDP			
	Log (y+1)	IH	Log (y+0.5)	0-1
AGOAccess	-0.108***	-0.112***	-0.127***	-0.030***
	(0.017)	(0.018)	(0.020)	(0.005)
AGOAccess × ER	0.202***	0.211***	0.240***	0.059***
	(0.046)	(0.048)	(0.054)	(0.012)
Observations	269497	269497	269497	269497

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log(fatalities +1) in column (1), the inverse hyperbolic sine transformation in column (2), log(fatalities +0.5) in column (3), and a binary variable that takes the value of 1 if the number of fatalities > 0 in column (4), where fatalities is based on data from UCDP. The unit of observation is the PRIO GRID cell. All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B9. AGOA and Luminosity: Different Diversity Measures

	(1)	(2)	(3)	(4)
AGOAccess	0.465***	0.459***	0.466***	0.471***
	(0.041)	(0.041)	(0.044)	(0.041)
AGOAccess × ER	-0.398***	-0.396***	-0.319***	-0.428***
	(0.096)	(0.092)	(0.086)	(0.103)
AGOAccess × Greenberg	0.149			
	(0.115)			
AGOAccess × ELF2		0.145		
		(0.093)		
AGOAccess × ELF9			-0.006	
			(0.063)	
AGOAccess × POL^{er}				0.597
				(0.383)
Observations	241072	241072	241072	241072

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (nighttime light + 1). The unit of observation is the PRIO GRID cell (resolution 0.5 × 0.5 decimal degrees, approximately 55km × 55km at the equator). All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. The sample includes 9,293 grid-cells spread across 48 sub-Saharan African countries for the period of 1992–2017. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B10. AGOA and Luminosity: Alternative Transformations of Dependent Variable

	Income Proxied by Nighlight			
	Log (y+1)	IH	Log (y+0.5)	0-1
AGOAccess	0.464***	0.466***	0.507***	0.131***
	(0.041)	(0.041)	(0.046)	(0.010)
AGOAccess × ER	-0.321***	-0.321***	-0.342***	-0.073***
	(0.085)	(0.086)	(0.096)	(0.020)
Observations	241072	241072	241072	269497

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log(nighttime light + 1) in column (1), the inverse hyperbolic sine transformation in column (2), log(nighttime light +0.5) in column (3), and a binary variable that takes the value of 1 if nighttime light > 0 in column (4). The unit of observation is the PRIO GRID cell (resolution 0.5 × 0.5 decimal degrees, approximately 55km × 55km at the equator). All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. The sample includes 8,670 grid-cells spread across 48 sub-Saharan African countries for the period of 1992–2017. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2. Ethnic Remoteness Reduces the Peace Dividend from Trade Access

Split by Manufacturing, Crops, Minerals or Oil, and Textiles

Table B11. AGOA and Conflict (UCDP): Split by Products

	AGOAccess		X World		X US		Suitability	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AGOA (Crops)	-0.037***	-0.101***	-0.040***	-0.110***	-0.014	-0.091***	-0.015	-0.051***
	(0.009)	(0.018)	(0.009)	(0.019)	(0.011)	(0.024)	(0.010)	(0.018)
AGOA (Manufacturing)	-0.061**	-0.040	-0.061**	-0.038	-0.045	0.035	-0.067**	-0.067
	(0.028)	(0.041)	(0.028)	(0.041)	(0.031)	(0.045)	(0.029)	(0.041)
AGOA (Minerals/Oil)	-0.052***	-0.040**	-0.063***	-0.042*	-0.072***	-0.065**	-0.054***	-0.050***
	(0.016)	(0.019)	(0.020)	(0.023)	(0.024)	(0.032)	(0.016)	(0.019)
AGOA (Apparel)	-0.065*	0.008	-0.064*	0.009	-0.091***	-0.048	-0.066**	-0.015
	(0.034)	(0.048)	(0.034)	(0.049)	(0.034)	(0.049)	(0.033)	(0.048)
AGOA (Crops) × ER		0.208***		0.224***		0.288***		0.113**
		(0.051)		(0.052)		(0.080)		(0.052)
AGOA (Manufacturing) × ER		-0.068		-0.076		-0.239**		0.001
		(0.088)		(0.088)		(0.094)		(0.087)
AGOA (Minerals/Oil) × ER		-0.030		-0.063		0.006		-0.010
		(0.061)		(0.078)		(0.092)		(0.061)
AGOA (Apparel) × ER		-0.187		-0.182		-0.097		-0.145
		(0.119)		(0.119)		(0.119)		(0.119)
Observations	269497	269497	269497	269497	269497	269497	269497	269497

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (fatalities + 1), where fatalities is based on data from UCDP. The different columns split *Production* between crops, manufacturing, minerals and oil, and apparel (including textiles). Columns 1 and 2 use our main definition of AGOAccess used in equation (2.1). Columns 3, 4, 5, and 6 use a definition of AGOAccess taking into account if the country has export capacity in eligible AGOA goods as defined in equation (2.3) to the US and to the rest of the world. Columns 7 and 8 measure AGOAccess considering if a location's land is suitable for AGOA-eligible crop as defined in equation (2.4). The unit of observation is the PRIO GRID cell. All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B12. AGOA and Luminosity: Split by Products

	AGOAccess		X World		X US		Suitability	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AGOA (Crops)	0.212***	0.303***	0.322***	0.392***	0.211***	0.249***	0.212***	0.303***
	(0.032)	(0.046)	(0.030)	(0.046)	(0.044)	(0.060)	(0.032)	(0.046)
AGOA (Manufacturing)	-0.045	0.117	-0.093	0.066	-0.135**	0.016	-0.045	0.117
	(0.057)	(0.091)	(0.058)	(0.092)	(0.062)	(0.101)	(0.057)	(0.091)
AGOA (Minerals/Oil)	0.260***	0.358***	0.199***	0.322***	0.184**	0.321**	0.260***	0.358***
	(0.055)	(0.083)	(0.064)	(0.100)	(0.077)	(0.127)	(0.055)	(0.083)
AGOA (Apparel)	-0.065	-0.242**	-0.047	-0.189*	0.020	-0.094	-0.065	-0.242**
	(0.069)	(0.116)	(0.068)	(0.111)	(0.067)	(0.112)	(0.069)	(0.116)
AGOA (Crops) × ER		-0.307***		-0.223**		-0.153		-0.307***
		(0.092)		(0.095)		(0.126)		(0.092)
AGOA (Manufacturing) × ER		-0.539***		-0.541***		-0.478**		-0.539***
		(0.200)		(0.203)		(0.220)		(0.200)
AGOA (Minerals/Oil) × ER		-0.406*		-0.481*		-0.493		-0.406*
		(0.236)		(0.276)		(0.368)		(0.236)
AGOA (Apparel) × ER		0.530**		0.416*		0.341		0.530**
		(0.229)		(0.226)		(0.235)		(0.229)
Observations	241072	241072	241072	241072	241072	241072	241072	241072

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (nighttime light + 1). Columns 1 and 2 use our main definition of AGOAccess used in equation (1) splitting *Production* between crops, manufacturing, minerals and oil, and apparel (including textiles). Columns 3, 4, 5, and 6 use a definition of AGOAccess taking into account if the country has export capacity in AGOA-eligible goods as defined in equation (3) to the US and to the rest of the world, splitting *Production* between crops, manufacturing, minerals and oil, and apparel (including textiles). Columns 7 and 8 measure AGOAccess considering if a location's land is suitable for AGOA-eligible crop as defined in equation (4), splitting *Production* between the same aforementioned sectors. The unit of observation is the PRIO GRID cell. All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

GSP vs. AGOA

Table B13. AGOA vs. GSP and Conflict (UCDP)

	(1)	(2)	(3)	(4)	(5)
AGOAccess (No GSP)	-0.041***		-0.122***		-0.122***
	(0.011)		(0.028)		(0.028)
GSPAccess		0.001		-0.012	-0.011
		(0.012)		(0.018)	(0.018)
AGOAccess × ER			0.241***		0.240***
			(0.081)		(0.081)
GSPAccess × ER				0.046	0.042
				(0.042)	(0.042)
Observations	269497	269497	269497	269497	269497

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (fatalities + 1), where fatalities is based on data from UCDP. GSPAcess represents the shock experienced in cells belonging to countries that benefited from GSP provisions in 1997 for being least developed countries (LDCs). AGOAccess represents the pure AGOA shock experienced by cells from the year 2001 onwards. The unit of observation is the PRIO GRID cell. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B14. AGOA vs. GSP and Luminosity

	(1)	(2)	(3)	(4)	(5)
AGOAccess (No GSP)	0.272***		0.403***		0.400***
	(0.049)		(0.064)		(0.064)
GSPAccess		0.347***		0.351***	0.349***
		(0.031)		(0.049)	(0.049)
AGOAccess × ER			-0.399***		-0.402***
			(0.121)		(0.121)
GSPAccess × ER				-0.015	-0.012
				(0.110)	(0.110)
Observations	241072	241072	241072	241072	241072

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (fatalities + 1), where fatalities is based on data from UCDP. GSPAcess represents the shock experienced in cells belonging to countries that benefited from GSP provisions in 1997 for being least developed countries (LDCs). AGOAccess represents the pure AGOA shock experienced by cells from the year 2001 onwards. The unit of observation is the PRIO GRID cell.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2. Ethnic Remoteness Reduces the Peace Dividend from Trade Access

ACLED Definition of Conflict

Table B15. AGOA and Conflict: Ethnic Remoteness (ACLED)

	Intensity of Conflict from ACLED						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
AGOAccess	-0.032***	-0.093***	-0.103***	-0.101***	-0.092***	-0.089***	-0.083***
	(0.012)	(0.021)	(0.022)	(0.022)	(0.020)	(0.033)	(0.022)
AGOAccess × ER		0.202***	0.195***	0.196***	0.203***	0.202***	0.198***
		(0.053)	(0.054)	(0.055)	(0.058)	(0.052)	(0.053)
AGOAccess × ELF			0.028				
			(0.026)				
AGOAccess × POL				0.083			
				(0.099)			
AGOAccess × Specialization					-0.008		
					(0.052)		
AGOAccess × Kinship						-0.009	
						(0.060)	
AGOAccess × Segmented							-0.016
							(0.016)
Observations	195153	195153	195153	195153	195153	195153	195153

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (fatalities + 1), where fatalities is based on data from ACLED. The unit of observation is the PRIO GRID cell (resolution 0.5 × 0.5 decimal degrees, approximately 55km × 55km at the equator). All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. The sample includes 9,293 grid-cells spread across 48 sub-Saharan African countries for the period of 1997–2017. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B16. AGOA and Conflict: Alternative Transformations of Dependent Variable (ACLED)

	(1)	(2)	(3)	(4)
AGOAccess	-0.093***	-0.094***	-0.105***	0.013***
	(0.021)	(0.022)	(0.024)	(0.005)
AGOAccess × ER	0.202***	0.205***	0.230***	0.010
	(0.053)	(0.054)	(0.061)	(0.013)
Observations	195153	195153	195153	269497

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log(fatalities +1) in column (1), the inverse hyperbolic sine transformation in column (2), log(fatalities +0.5) in column (3), and a binary variable that takes the value of 1 if the number of fatalities > 0 in column (4), where fatalities is based on data from ACLED. The unit of observation is the PRIO GRID cell. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B17. AGOA and Conflict: Different Diversity Measures (ACLED)

	(1)	(2)	(3)	(4)
AGOAccess	-0.093***	-0.093***	-0.094***	-0.095***
	(0.021)	(0.021)	(0.021)	(0.022)
AGOAccess × ER	0.203***	0.202***	0.200***	0.229***
	(0.072)	(0.063)	(0.056)	(0.078)
AGOAccess × Greenberg	-0.003			
	(0.064)			
AGOAccess × ELF2		0.000		
		(0.046)		
AGOAccess × ELF9			0.005	
			(0.031)	
AGOAccess × POL^{er}				-0.152
				(0.208)
Observations	195153	195153	195153	195153

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (fatalities + 1), where fatalities is based on data from ACLED. The unit of observation is the PRIO GRID cell. All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B18. AGOA and Conflict: Environmental Variables (ACLED)

	(1)	(2)	(3)	(4)
AGOAccess	-0.093***	-0.158***	-0.103***	-0.065***
	(0.021)	(0.034)	(0.022)	(0.022)
AGOAccess × ER	0.202***	0.189***	0.187***	0.183***
	(0.053)	(0.053)	(0.053)	(0.051)
AGOAccess × Crop Unsuitability		0.014**		
		(0.005)		
AGOAccess × Malaria Suitability			0.032***	
			(0.011)	
AGOAccess × Tsetse Suitability				-0.023***
				(0.008)
Observations	195153	195153	195153	195153

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (fatalities + 1), where fatalities is based on data from ACLED. The unit of observation is the PRIO GRID cell (resolution 0.5 × 0.5 decimal degrees, approximately 55km × 55km at the equator). All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. The sample includes 9,293 grid-cells spread across 48 sub-Saharan African countries for the period of 1989–2017. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B19. AGOA and Conflict: Alternative Definitions of AGOA Exposure (ACLED)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AGOAGeo	-0.086**	-0.111**						
	(0.041)	(0.044)						
AGOAGeo × ER		0.110**						
		(0.043)						
AGOAExp (World)			-0.039***	-0.110***				
			(0.012)	(0.022)				
AGOExp (World) × ER				0.233***				
				(0.054)				
AGOAExp (US)					-0.040***	-0.100***		
					(0.015)	(0.025)		
AGOAExp (US) × ER						0.234***		
						(0.076)		
AGOASuit							-0.002	-0.038*
							(0.013)	(0.023)
AGOASuit × ER								0.116**
								(0.054)
Observations	195153	195153	195153	195153	195153	195153	195153	195153

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (fatalities + 1), where fatalities is based on data from ACLED. Columns (1) and (2) use the broad definition of AGOA exposure without requiring the production of AGOA-eligible goods as defined in equation (2.2). Columns (3), (4), (5), and (6) use a narrow definition of AGOA that takes into account if the country has export capacity in eligible AGOA goods to either the rest of the world or the U.S. as defined in equation (2.3). Columns (7) and (8) measure make AGOA exposure conditional on a location's land being suitable for AGOA-eligible crops as defined in equation (2.4). The unit of observation is the PRIO GRID cell. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B20. AGOA and Conflict (ACLED): Split by Products

	AGOAccess		X World		X US		Suitability	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AGOA (Crops)	-0.026**	-0.088***	-0.035***	-0.110***	0.002	-0.057**	0.014	-0.016
	(0.012)	(0.022)	(0.012)	(0.022)	(0.013)	(0.026)	(0.015)	(0.024)
AGOA (Manufacturing)	-0.043	-0.039	-0.043	-0.036	-0.017	0.068	-0.049	-0.066
	(0.039)	(0.061)	(0.039)	(0.061)	(0.041)	(0.066)	(0.039)	(0.061)
AGOA (Minerals/Oil)	-0.039*	-0.043	-0.055**	-0.057	-0.083***	-0.131**	-0.041**	-0.053*
	(0.021)	(0.030)	(0.025)	(0.040)	(0.032)	(0.054)	(0.021)	(0.031)
AGOA (Apparel)	0.011	0.061	0.014	0.070	-0.018	-0.028	0.005	0.026
	(0.043)	(0.070)	(0.043)	(0.070)	(0.044)	(0.076)	(0.043)	(0.069)
AGOA (Crops) × ER		0.207***		0.243***		0.228**		0.099
		(0.058)		(0.058)		(0.094)		(0.062)
AGOA (Manufacturing) × ER		-0.013		-0.023		-0.271**		0.061
		(0.133)		(0.132)		(0.135)		(0.130)
AGOA (Minerals/Oil) × ER		0.032		0.022		0.193		0.054
		(0.079)		(0.102)		(0.125)		(0.079)
AGOA (Apparel) × ER		-0.124		-0.134		0.060		-0.064
		(0.143)		(0.143)		(0.156)		(0.141)
Observations	195153	195153	195153	195153	195153	195153	195153	195153

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is log (fatalities + 1), where fatalities is based on data from ACLED. Columns 1 and 2 use our main definition of AGOAccess used in equation (2.1), splitting *Production* between crops, manufacturing, minerals and oil, and apparel (including textiles). Columns 3, 4, 5, and 6 use a definition of AGOAccess taking into account if the country has export capacity in AGOA-eligible goods as defined in equation (2.3) to the US and to the rest of the world, splitting *Production* between crops, manufacturing, minerals and oil, and apparel (including textiles). Columns 7 and 8 measure AGOAccess considering if a location's land is suitable for AGOA-eligible crops as defined in equation (2.4), splitting *Production* between crops, manufacturing, minerals and oil, and apparel (including textiles). The unit of observation is the PRIO GRID cell (resolution 0.5 × 0.5 decimal degrees, approximately 55km × 55km at the equator). All specifications control for rainfall deviation, temperature deviation, cell FEs, and country-specific year FEs. The sample includes 9,293 grid-cells spread across 48 sub-Saharan African countries for the period of 1997–2017. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.B.3 Robustness: Individual Level Data

Food Poverty: Baseline Definition of AGOA

Table B21. AGOA and Food Poverty – Remoteness from the Dominant Group

	(1)	(2)	(3)	(4)	(5)	(6)
AGOAccess	-0.077*	-0.078*	-0.078*	-0.113*	-0.115*	-0.117*
	(0.047)	(0.045)	(0.047)	(0.064)	(0.064)	(0.064)
AGOAccess × Indiv ER^{dom}	0.089***		0.088***	0.116***		0.105***
	(0.028)		(0.031)	(0.027)		(0.032)
AGOAccess × Cell ER^{dom}		0.063*	0.004		0.100***	0.033
		(0.037)	(0.046)		(0.036)	(0.049)
Individual Controls						
Observations	114176	114176	114176	72112	72112	72112

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is based on the answer to the question: "Over the past year, how often, if ever, have you or your family gone without: Enough food to eat?". It is coded as 0 (if answer is never) or 1 (if answer is sometimes/several times/frequently/many times/always). The unit of observation is the individual. The sample is based on six rounds of the Afrobarometer surveys conducted between 1999–2015 comprising approximately between 17k and 22k individual per round spread across 12 countries (see Appendix 2.A.3 for full list of countries). Indiv ER refers to the individual ethnic remoteness from the dominant group in the country. Cell ER refers to cell-level ethnic remoteness of the cell in which the individual resides. All regressions control for rainfall and temperature shocks, country-specific cell FE, country-specific year FE and individual ethnolinguistic group FE. Columns (4)-(6) include additional individual controls for professions, age bracket, gender, and rural location. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B22. AGOA and Food Poverty: Additional Cell Controls

	(1)	(2)	(3)	(4)	(5)	(6)
AGOAccess	-0.113**	-0.111**	-0.115**	-0.105*	-0.109*	-0.128**
	(0.055)	(0.056)	(0.056)	(0.054)	(0.066)	(0.055)
AGOAccess × Indiv ER	0.167***	0.168***	0.168***	0.168***	0.168***	0.169***
	(0.050)	(0.050)	(0.049)	(0.049)	(0.050)	(0.049)
AGOAccess × Cell ER	0.016	0.019	0.014	0.029	0.015	0.028
	(0.077)	(0.077)	(0.076)	(0.080)	(0.075)	(0.077)
AGOAccess × Cell ELF		-0.007				
		(0.038)				
AGOAccess × Cell POL			0.017			
			(0.129)			
AGOAccess × Cell Specialization				-0.044		
				(0.067)		
AGOAccess × Cell Kinship					-0.012	
					(0.100)	
AGOAccess × Cell Segmented						0.036
						(0.032)
Observations	114176	114176	114176	114176	114176	114176

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is based on the answer to the question: "Over the past year, how often, if ever, have you or your family gone without: Enough food to eat?". It is coded as 0 (if answer is never) or 1 (if answer is sometimes/several times/frequently/many times/always). The unit of observation is the individual. The sample is based on six rounds of the Afrobarometer surveys conducted between 1999–2015 comprising approximately between 17k and 22k individual per round spread across 12 countries (see Appendix 2.A.3 for full list of countries). Indiv ER refers to the individual ethnic remoteness from her fellow citizens in the country. Cell ER refers to cell-level ethnic remoteness of the cell in which the individual resides. All regressions control for rainfall and temperature shocks, country-specific cell FE, country-specific year FE and individual ethnolinguistic group FE. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B23. AGOA and Food Poverty: Alternative Diversity controls

	(1)	(2)	(3)	(4)	(5)
AGOAccess	-0.113**	-0.116**	-0.118**	-0.109**	-0.124**
	(0.055)	(0.053)	(0.053)	(0.054)	(0.054)
AGOAccess × Indiv ER	0.167***	0.168***	0.169***	0.168***	0.165***
	(0.050)	(0.049)	(0.049)	(0.050)	(0.049)
AGOAccess × Cell ER	0.016	0.113	0.118	0.029	0.143**
	(0.077)	(0.075)	(0.075)	(0.076)	(0.070)
AGOAccess × Cell Greenberg		-0.129*			
		(0.068)			
AGOAccess × Cell ELF2			-0.133*		
			(0.069)		
AGOAccess × Cell ELF9				-0.029	
				(0.045)	
AGOAccess × Cell POL^{er}					-0.421**
					(0.198)
Observations	114176	114176	114176	114176	114176

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is based on the answer to the question: "Over the past year, how often, if ever, have you or your family gone without: Enough food to eat?". It is coded as 0 (if answer is never) or 1 (if answer is sometimes/several times/frequently/many times/always). The unit of observation is the individual. The sample is based on six rounds of the Afrobarometer surveys conducted between 1999–2015 comprising approximately between 17k and 22k individual per round spread across 12 countries (see Appendix 2.A.3 for full list of countries). Indiv ER refers to the individual ethnic remoteness from her fellow citizens in the country. Cell ER refers to cell-level ethnic remoteness of the cell in which the individual resides. All regressions control for rainfall and temperature shocks, country-specific cell FE, country-specific year FE and individual ethnolinguistic group FE. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B24. AGOA and Food Poverty: Environmental Controls

	(1)	(2)	(3)	(4)
AGOAccess	-0.113**	-0.034	-0.107*	-0.102*
	(0.055)	(0.069)	(0.054)	(0.052)
AGOAccess × Indiv ER	0.167***	0.168***	0.184***	0.163***
	(0.050)	(0.049)	(0.049)	(0.049)
AGOAccess × Cell ER	0.016	0.024	0.030	-0.003
	(0.077)	(0.077)	(0.067)	(0.083)
AGOAccess × Crop Unsuitability		-0.015*		
		(0.009)		
AGOAccess × Malaria Suitability			0.043	
			(0.031)	
AGOAccess × Tsetse Suitability				-0.009
				(0.015)
Observations	114176	114176	114176	114176

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is based on the answer to the question: "Over the past year, how often, if ever, have you or your family gone without: Enough food to eat?". It is coded as 0 (if answer is never) or 1 (if answer is sometimes/several times/frequently/many times/always). The unit of observation is the individual. The sample is based on six rounds of the Afrobarometer surveys conducted between 1999–2015 comprising approximately between 17k and 22k individual per round spread across 12 countries (see Appendix 2.A.3 for full list of countries). Indiv ER refers to the individual ethnic remoteness from her fellow citizens in the country. All regressions control for rainfall and temperature shocks, country-specific cell FE, country-specific year FE and individual ethnolinguistic group FE. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B25. AGOA and Income Poverty: Additional Cell Controls (Dominant)

	(1)	(2)	(3)	(4)	(5)	(6)
AGOAccess	-0.061	-0.048	-0.041	-0.057	-0.050	-0.059
	(0.072)	(0.074)	(0.075)	(0.074)	(0.084)	(0.073)
AGOAccess × Indiv ER^{dom}	0.088**	0.087**	0.086**	0.088**	0.088**	0.088**
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
AGOAccess × Cell ER^{dom}	0.109*	0.134**	0.143**	0.123	0.113*	0.106*
	(0.059)	(0.065)	(0.063)	(0.080)	(0.062)	(0.061)
AGOAccess × Cell ELF		-0.043				
		(0.052)				
AGOAccess × Cell POL			-0.197			
			(0.149)			
AGOAccess × Cell Specialization				-0.027		
				(0.103)		
AGOAccess × Cell Kinship					-0.030	
					(0.134)	
AGOAccess × Cell Segmented						-0.007
						(0.041)
Observations	108463	108463	108463	108463	108463	108463

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is based on the answer to the question: "Over the past year, how often, if ever, have you or your family gone without: A cash income?". It is coded as 0 (if answer is never) or 1 (if answer is sometimes/several times/frequently/many times/always). The unit of observation is the individual. The sample is based on six rounds of the Afrobarometer surveys conducted between 1999–2015 comprising approximately between 13k and 22k individuals per round spread across 12 countries (see Appendix 2.A.3 for full list of countries). Indiv ER refers to the individual ethnic remoteness from the dominant group in the country. All regressions control for rainfall and temperature shocks, country-specific cell FE, country-specific year FE and individual ethnolinguistic group FE. sym* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B26. AGOA and Income Poverty: Alternative Diversity Indices (Dominant)

	(1)	(2)	(3)	(4)	(5)
AGOAccess	-0.061	-0.059	-0.060	-0.060	-0.057
	(0.072)	(0.073)	(0.073)	(0.073)	(0.073)
AGOAccess × Indiv $\mathrm{ER}^{\mathrm{dom}}$	0.088^{**}	0.088^{**}	0.088^{**}	0.088^{**}	0.087^{**}
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
AGOAccess × Cell $\mathrm{ER}^{\mathrm{dom}}$	0.109^{*}	0.126^{*}	0.118^{*}	0.115	0.153^{**}
	(0.059)	(0.068)	(0.067)	(0.073)	(0.070)
AGOAccess × Cell Greenberg		-0.032			
		(0.102)			
AGOAccess × Cell ELF2			-0.016		
			(0.101)		
AGOAccess × Cell ELF9				-0.010	
				(0.078)	
AGOAccess × Cell POL^{er}					-0.229
					(0.325)
Observations	108463	108463	108463	108463	108463

Notes: Standard errors in parentheses corrected for spatial correlation within a 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is based on the answer to the question: "Over the past year, how often, if ever, have you or your family gone without: Enough food to eat?". It is coded as 0 (if answer is never) or 1 (if answer is sometimes/several times/frequently/many times/always). The unit of observation is the individual. The sample is based on six rounds of the Afrobarometer surveys conducted between 1999–2015 comprising approximately between 17k and 22k individual per round spread across 12 countries (see Appendix 2.A.3 for full list of countries). Indiv ER refers to the individual ethnic remoteness from the dominant group in the country. Cell ER refers to cell-level ethnic remoteness of the cell in which the individual resides. All regressions control for rainfall and temperature shocks, country-specific cell FE, country-specific year FE and individual ethnolinguistic group FE. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.C Sector-Wise Matching to AGOA Tariff Lines

2.C.1 Crops Matching

Crop FAO	Description FAO	HSTS-8 US Tariff Line	Decision
alfalfa	Alfalfa.	Alfalfa (lucerne) meal and pellets.	1
almond	Almonds, with shell.	Almonds, fresh or dried, in shell.	1
apple	Apples.	Apples, otherwise prepared or preserved, nesi.	1
apricot	Apricots.	Apricots, fresh.	1
artichoke	Artichokes.	Artichokes, prepared or preserved otherwise than by vinegar or acetic acid, not frozen.	1
asparagus	Asparagus.	Asparagus, nesi, fresh or chilled.	1
avocado	Avocados.	Avocados, fresh or dried.	1
bambara	Bambara beans, bambara groundnut, earth pea.	Beans nesi, fresh or chilled, shelled or unshelled.	1
barley	Barley, two-row barley, six-row barley, four-row barley.	Barley, other than for malting purposes.	1
bean	Beans, dry; kidney/haricot bean, lima/butter bean, adzuki bean, mungo bean, black gram, scarlet runner bean, rice bean, moth bean, and tepary bean.	Beans nesi, fresh or chilled, shelled or unshelled.	1
berrynes	Berries nes; blackbeery, loganberry, white/red mulberry, myrtle berry huckleberry, dangleberry, and other berries not separately identified.	Boysenberries, frozen, in water or containing added sweetening.	0
blueberry	Blueberries, wild bilberry, whortleberry, american blueberry.	Blueberries, otherwise prepared or preserved, nesi.	1
broadbean	Broad beans, horse beans, broad bean, and field bean.	Beans nesi, uncooked or cooked by steaming or boiling in water, frozen, reduced in size.	1
cabbage	Cabbages and other brassicas chinese, mustard cabbage, pak-choi, with/red/savoy cabbage, brussels sprouts, collards, kale, and kohlarabi.	Kohlrabi, kale and similar edible brassicas nesi, including sprouting broccoli, fresh or chilled.	1
carrots	Carrots and turnips.	Carrots, fresh or chilled, reduced in size.	1
cauliflower	Cauliflowers and broccoli brassica.	Cauliflower seeds of a kind used for sowing.	1

(continued on next page)

Crop FAO	Description FAO	HSTS-8 US Tariff Line	Decision
cerealnes	Cereals, nes, including canagua or coaihua, quihuicha or inca wheat, adlay or job's tears, wild rice and other cereal crops that are not identified separately.	Cereals nesi (including wild rice).	1
cherry	Cherries, sweet cherry, hard-fleshed cherry and hear cherry.	Cherries, dried.	1
chestnut	Chestnut.	Acorns and horse-chestnuts, of a kind used in animal feeding.	0
chicory	Chicory roots, unroasted chicory roots.	Roasted chicory and other roasted coffee substitutes and extracts, essences and concentrates thereof.	1
citrusnes	Fruit, citrus nes; bergamot, citron, chinotto, kumquat and some minor varieties of citrus.	Citrus fruit or peel of citrus or other fruit, except mixtures, preserved by sugar (drained, glace or crystallized).	1
clover	Clover.	White and ladino clover seed of a kind used for sowing.	1
cocoa	Cocoa, beans, including whole or broken, raw or roasted.	Chocolate/oth preps with cocoa, ov 2kg but n/o 4.5 kg, n/o 65% by wt of sugar, not in blocks 4.5 kg or more, subj to GN 15.	0
coffee	Coffee, green (arabica, robusta, liberica). Raw coffee in all forms.	Coffee substitutes containing coffee.	0
cotton	Seed cotton; unginned cotton, cottonseed, cotton lint and linters.	Cotton seeds, whether or not broken.	1
cowpea	Cow peas; dry cowpea, blackeye pea/bean.	Black-eye cowpeas, shelled, prepared or preserved otherwise than by vinegar or acetic acid, not frozen.	1
cucumberetc	Cucumbers and gherkins.	Cucumbers, including gherkins, fresh or chilled, if entered May 1 to June 30, inclusive, or Sept. 1 to Nov. 30, inclusive, in any year.	1
currant	Currant black, red and white. Trade data may include gooseberries.	Currant and berry fruit jellies.	1
date	Dates, including fresh and dried fruit.	Dates, fresh or dried, other than whole.	1
fig	Figs.	Figs, fresh or dried, other than whole (including fig paste).	1
flax	Flax fibre and tow broken, scutched, hackled, etc, but not spun. Production in its raw state.	Flaxseed (linseed), whether or not broken.	0

(continued on next page)

Crop FAO	Description FAO	HSTS-8 US Tariff Line	Decision
fruitnes	Fruit, fresh nes; azarole, babaco, elderberry, jujube, litchi, loquat, medlar, pawpaw, pomegranate, prickly pear, rose hips, rowanberry, service-apple, tamarind, tree-strawberry, and other fresh fruit not identified separately.	Fruit nesi, dried, other than that of headings 0801 to 0806, and excluding mixtures.	0
grape	Grapes; table and wine grapes.	Grapes, dried, other than raisins.	1
grapefruit	Grapefruit (inc pomelos).	Grapefruit (other than peel or pulp), otherwise prepared or preserved, nesi.	1
greenbean	Beans, green.	Beans nesi, uncooked or cooked by steaming or boiling in water, frozen, reduced in size.	1
greenbroadbean	Vegetables, leguminous.	Leguminous vegetables nesi, uncooked or cooked by steaming or boiling in water, frozen.	1
greencorn	Maize, green, particularly var saccharata. Saccharata is known as sweet corn.	Sweet corn, uncooked or cooked by steaming or boiling in water, frozen.	1
greenonion	Onion, shallots, green; welsh onions, and young onions pulled before the bulb has enlarged.	Dried onion powder or flour.	0
groundnut	Groundnuts, with shell.	Peanuts (ground-nuts), not roasted or cooked, in shell, subject to add. US note 2 to Ch.12.	1
hazelnut	Hazelnuts, with shell.	Hazelnuts or filberts, fresh or dried, in shell.	1
hop	Hops; hop cones, fresh or dried, whether or not ground, powdered or in the form of pellets.	Saps and extracts of hops.	1
lemonlime	Lemons and limes.	Lemons, fresh or dried.	1
linseed	Linseed. An annual herbaceous that is cultivated for its fibre as well as its oil.	Linseed oil, crude, and its fractions, not chemically modified.	1
maize	Maize corn. A grain with a high germ content. Used largely for animal feed and commercial starch production.	Corn (maize) oil, crude, and its fractions, not chemically modified.	0
maizefor	Maizefor, forage.	Corn (maize) oil, crude, and its fractions, not chemically modified.	0
melonetc	Melons, other.	Other melons nesoi, fresh, if entered during the period from June 1 through November 30, inclusive.	1

(continued on next page)

Crop FAO	Description FAO	HSTS-8 US Tariff Line	Decision
melonseed	Melonseed.	Other melons nesoi, fresh, if entered during the period from June 1 through November 30, inclusive.	0
millet	Millet; japanese millet, african millet, proso millet, ditch millet, pearl or cattail millet and foxtail millet.	Millet.	1
mixedgrain	Grain, mixed. A mixture of cereal species that are sown and harvested together. The mixture wheat/rye is known as meslin, but in trade is usually classified with wheat.	Grains of barley, hulled, pearled, clipped, sliced, kibbled or otherwise worked, but not rolled or flaked.	0
mixedgrass	Mixedgrass, forage.	Rye grass seed of a kind used for sowing.	0
mushrooms	Mushrooms and truffles.	Mushrooms, fresh or chilled.	1
mustard	Mustard seed; white and black mustard.	Rapeseed, colza or mustard oil, crude, and their fractions, not chemically modified, nesi.	1
nutnes	Nuts, nes; pecan, butter or swarri nut, pili nut, java almond, chinese olives, paradise or sapucaia, queensland, macadamia nut, pignolia nut, and other nuts that are not identified separately.	Nuts nesi, fresh or dried, shelled.	1
oilseedfor	Oilseedfor, forage.	Flours and meals of oil seeds or oleaginous fruits other than those of mustard or soybeans.	0
oilseednes	Oilseed, nes; beech nut, and other oilseeds, oleaginous fruits, and nuts that are not identified separately.	Flours and meals of oil seeds or oleaginous fruits other than those of mustard or soybeans.	1
olive	Olives; table olives and olives for oil.	Olives, fresh or chilled.	1
onion	Onions, dry; onions at a mature stage, but not dehydrated onions.	Dried onions whole, cut, sliced or broken, but not further prepared.	1
orange	Oranges common, sweet orange, and bitter orange.	Orange juice, fortified with vitamins or minerals.	1
papaya	Papayas.	Papayas, frozen, in water or containing added sweetening.	1
peachetc	Peaches and nectarines.	Peaches, including nectarines, fresh, if entered during the period from June 1 through November 30, inclusive.	1
pear	Pears.	Pears and quinces, fresh, if entered during the period from July 1 through the following March 31, inclusive.	1
pineapple	Pineapples ananas.	Pineapples, fresh or dried, not reduced in size, in bulk.	1

(continued on next page)

Crop FAO	Description FAO	HSTS-8 US Tariff Line	Decision
plum	Plums and sloes.	Plums, prunes and sloes, fresh, if entered during the period from June 1 through December 31, inclusive.	1
potato	Potatoes.	Potatoes, uncooked or cooked by steaming or boiling in water, frozen.	1
pyrethrum	Pyrethrum; leaves, stems and flowers. For insecticides, fungicides and similar products.	Insecticides, nesoi, for retail sale or as preparations or articles.	0
quince	Quinces.	Pears and quinces, fresh, if entered during the period from July 1 through the following March 31, inclusive.	1
rapeseed	Rapeseed Brassica. Valued mainly for its oil.	Rapeseed, colza or mustard oil, other than crude, and their fractions, whether or not refined, not chemically modified, nesi.	1
rasberry	Raspberries.	Raspberries and loganberries, fresh, if entered during the period from September 1 through the following June 30, inclusive.	1
rice	Rice, paddy. Also known as rice in the husk and rough rice.	Rice in the husk (paddy or rough).	1
rootnes	Roots and tuber, nes; arracha, arrowroot, chuga, sago palm, oca and ullucu, yam bean, jicama, mashua, Jerusalem artichoke, topinambur, and other tubers roots, or rhizomes, fresh that are not identified separately.	Fresh or chilled arrowroot, salep, Jerusalem artichokes and similar roots and tubers nesoi, whether or not sliced or in the form of pellets.	1
rubber	Rubber, latex. The liquid secreted by the rubber tree.	Sports footwear w/outer soles and uppers of rubber or plastics, nesi, valued over 3*butnotover*6.50/pair.	0
rye	Rye cereale; mainly used in making bread.	Rye flour.	1
ryefor	Ryefor, forage.	Rye grass seed of a kind used for sowing.	1
safflower	Safflower. Valued mainly for its oil.	Sunflower seed or safflower oil, other than crude, and their fractions, whether or not refined, but not chemically modified.	1
soybean	Soybeans; oil crop.	Soybean oil, other than crude, and its fractions, whether or not refined, but not chemically modified, nesi.	1
spinach	Spinach; New Zealand spinach, and orage spinach.	Spinach, New Zealand spinach and orache spinach (garden spinach), fresh or chilled.	1

(continued on next page)

Crop FAO	Description FAO	HSTS-8 US Tariff Line	Decision
stonefruits	Fruit, stone nes; other stone fruit not separately identified. For some countries, apricots, cherries, peachers, nectarines, and plums.	Apricots, fresh.	0
strawberry	Strawberries.	Strawberries, otherwise prepared or preserved, nesi.	1
stringbean	String beans.	Beans nesi, fresh or chilled, shelled or unshelled.	1
sugarbeet	Sugar beet.	Sugar beet, fresh, chilled, frozen or dried, whether or not ground.	1
sugarcane	Sugar cane.	Sugar confectionery nesoi, w/o cocoa, dairy products subject to add. US note 1 to chap. 4: subject to add US note 10 to chapter 4.	0
sugarnes	Sugar crops, nes; sugar maple, sweet sorghum, sugar palm and other minor sugar crops of local importance.	Sugar confectionery nesoi, w/o cocoa, dairy products subject to add. US note 1 to chap. 4: subject to add US note 10 to chapter 4.	0
sunflower	Sunflower. Valued mainly for its oil.	Sunflower-seed or safflower oil, crude, and their fractions, whether or not refined, not chemically modified.	1
sweetpotatos	Sweet potatos.	Potatoes, uncooked or cooked by steaming or boiling in water, frozen.	0
tangetc	Tangerines, mandarins, clementines, and satsumas mandarin.	Mandarins (including tangerines and satsumas); clementines, wilkings and similar citrus hybrids, fresh or dried.	1
tobacco	Tobacco, unmanufactured; unmanufactures dry tobacco, including refuse that is not stemmed or stripped, or is partly or wholly stemmed or strippied.	Tobacco, partly or wholly stemmed/stripped, n/threshed or similarly proc., not or n/over 35% wrapper, flue-cured burley etc, not for cigaret.	1
tomato	Tomatoes.	Tomatoes, whole or in pieces, prepared or preserved otherwise than by vinegar or acetic acid.	1

(continued on next page)

Crop FAO	Description FAO	HSTS-8 US Tariff Line	Decision
vegetablenes	Vegetables, fresh nes; bamboo shoots, beets/chards, capers, cardoons, celery, chervil, cress, fennel, horseadish, marjoram, sweet, oyster plant, parsley, parsnips, radish, rhubarb, rutabagas/swedes, savory, scorzonera, sorrel, soybean sprouts tarragon, watercress, and other vegetables that are not separately identified because of their minor relevance.	Vegetables, nesoi, fresh or chilled.	1
walnut	Walnuts, with shell.	Walnuts, fresh or dried, shelled.	1
watermelon	Watermelons.	Watermelons, fresh, if entered during the period April 1 through November 30, inclusive.	1
wheat	Wheat.	Wheat or meslin flour.	1

Notes: This table contains the complete set of crops that appear in at least one AGOA-eligible tariff line for the year 2000. Only crops that are listed as the main component of the tariff line are included in our analysis. Column "Crop FAO" lists the crop names as given by FAO. Column "Description FAO" provides FAO's description of each crop. Column "HSTS-8 US Tariff Line" shows the exact name of one tariff line where the FAO crop appears. Column "Decision" takes a value of 1 if we classify the crop as AGOA-eligible. We classify a crop as AGOA-eligible if it is the primary product in at least one tariff line.

2.C.2 Mining Matching

Minerals	HSTS-8 US Tariff Line	Decision
Bauxite	Aluminum alloys, w/25% or more by weight of silicon, unwrought nesoi	1
Aluminum	Aluminum alloys, w/25% or more by weight of silicon, unwrought nesoi	1
Copper	Copper, containers a kind normally carried on the person, in the pocket or in the handbag	0
Gold	Watch cases of gold- or silver-plated base metal	0
Iron Ore	Iron/nonalloy steel, width 600mm+, flat-rolled products, clad	1
Lead	Iron/nonalloy steel, width 600mm+, flat-rolled products, plated or coated with lead, including terneplate	0
Silver	Silver ores and concentrates	1
Tin	Iron/nonalloy steel, width less th/600mm, flat-rolled products, plated or coated with tin	0
Zinc	Zinc (o/than alloy), unwrought, casting-grade zinc, containing at least 97.5% but less than 99.99% by weight of zinc	1
Cobalt	Cobalt alloy, unwrought	1

(continued on next page)

Minerals	HSTS-8 US Tariff Line	Decision
Ferromanganese	Ferromanganese containing by weight more than 4 percent of carbon	1
Manganese	Ferromanganese containing by weight more than 4 percent of carbon	1
Niobium	Niobium (columbium), unwrought; niobium, powders	1
Tungsten	Tungsten, unwrought (including bars and rods obtained simply by sintering)	1
Uranium	Alloys, dispersions (including cermets), ceramic products and mixtures containing natural uranium or natural uranium compounds	0
Vanadium	Vanadium (o/than waste & scrap) and articles thereof	1
Zircon	Zirconium, unwrought; zirconium, powders	1
Molybdenum	Molybdenum ores and concentrates, roasted	1
Ferromolybdenum	Ferromolybdenum	1
Titanium	Titanium, unwrought; titanium, powders	1
Scandium	Rare-earth metals, scandium and yttrium, whether or not intermixed or interalloyed	1
Yttrium	Rare-earth metals, scandium and yttrium, whether or not intermixed or interalloyed	1
Chromium	Ferrochromium containing by weight more than 3 percent but not more than 4 percent of carbon	1
Ferrochrome	Ferrochromium containing by weight more than 3 percent but not more than 4 percent of carbon	1

Notes: This table contains the complete set of minerals from the S&P Metals and Mining dataset that appear in at least one AGOA-eligible tariff line for the year 2000. Only minerals that are listed as the main component of the tariff line are included in our analysis. Column "Minerals" lists the minerals as shown in the Metals & Mining dataset. Column "HSTS-8 US Tariff Line" shows the exact name of an AGOA-eligible tariff line where the mineral appears. Column "Decision" is marked as 1 if we classify the mineral as AGOA-eligible. A mineral is classified as AGOA-eligible if it is the primary product in at least one tariff line.

2.C.3 Industries Matching

NAICS Code	Industry	Total Tariffs	AGOA Tariffs	Apparel Tariffs	Textiles Tariffs	Decision
1111	Oilseed and Grain Farming	90	18	-	4	No
1112	Vegetable and Melon Farming	113	31	-	4	Yes
1113	Fruit and Tree Nut Farming	81	34	-	-	Yes
1114	Greenhouse, Nursery, and Floriculture Production	40	8	-	-	No
1119	Other Crop Farming	233	34	-	16	No
1121	Cattle Ranching and Farming	6	1	-	-	No
1123	Poultry and Egg Production	7	6	-	-	Yes
1124	Sheep and Goat Farming	12	1	-	5	Yes
1125	Aquaculture	252	36	-	18	No
1129	Other Animal Production	26	5	-	3	No

(continued on next page)

NAICS Code	Industry	Total Tariffs	AGOA Tariffs	Apparel Tariffs	Textiles Tariffs	Decision
1132	Forest Nurseries and Gathering of Forest Products	30	3	-	-	No
1141	Fishing	98	5	-	-	No
2111	Oil and Gas Extraction	234	16	-	-	No
2122	Metal Ore Mining	26	4	-	-	No
3111	Animal Food Manufacturing	14	8	-	-	Yes
3112	Grain and Oilseed Milling	280	73	-	-	Yes
3113	Sugar and Confectionery Product Manufacturing	235	23	-	-	No
3114	Fruit and Vegetable Preserving and Specialty Food Manufacturing	373	117	-	-	Yes
3115	Dairy Product Manufacturing	302	189	-	-	Yes
3116	Animal Slaughtering and Processing	180	79	-	5	Yes
3117	Seafood Product Preparation and Packaging	102	23	-	-	No
3118	Bakeries and Tortilla Manufacturing	127	9	-	-	No
3119	Other Food Manufacturing	353	106	-	-	Yes
3121	Beverage Manufacturing	64	16	-	-	Yes
3122	Tobacco Manufacturing	19	12	-	-	Yes
3131	Fiber, Yarn, and Thread Mills	236	-	-	208	Yes
3132	Fabric Mills	546	-	-	454	Yes
3133	Textile and Fabric Finishing and Fabric Coating Mills	282	2	2	180	Yes
3141	Textile Furnishings Mills	111	-	-	79	Yes
3149	Other Textile Product Mills	957	13	405	182	Yes
3151	Apparel Knitting Mills	16	-	2	-	No
3152	Cut and Sew Apparel Manufacturing	671	12	405	29	Yes
3159	Apparel Accessories and Other Apparel Manufacturing	199	11	66	19	Yes
3161	Leather and Hide Tanning and Finishing	144	12	-	3	No
3162	Footwear Manufacturing	93	80	-	-	Yes
3169	Other Leather and Allied Product Manufacturing	83	32	-	-	Yes
3212	Veneer, Plywood, and Engineered Wood Product Manufacturing	59	1	-	-	No
3219	Other Wood Product Manufacturing	511	31	-	17	No
3221	Pulp, Paper, and Paperboard Mills	118	1	-	-	No
3222	Converted Paper Product Manufacturing	110	5	-	6	No

(continued on next page)

NAICS Code	Industry	Total Tariffs	AGOA Tariffs	Apparel Tariffs	Textiles Tariffs	Decision
3231	Printing and Related Support Activities	49	-	4	3	No
3241	Petroleum and Coal Products Manufacturing	304	212	-	-	Yes
3251	Basic Chemical Manufacturing	1292	356	-	-	Yes
3252	Resin, Synthetic Rubber, and Artificial and Synthetic Fibers and Filaments Manufacturing	153	-	-	33	No
3253	Pesticide, Fertilizer, and Other Agricultural Chemical Manufacturing	50	4	-	-	No
3254	Pharmaceutical and Medicine Manufacturing	159	33	-	-	No
3255	Paint, Coating, and Adhesive Manufacturing	103	30	-	-	Yes
3256	Soap, Cleaning Compound, and Toilet Preparation Manufacturing	73	8	-	-	No
3259	Other Chemical Product and Preparation Manufacturing	1257	262	-	3	No
3261	Plastics Product Manufacturing	272	13	-	14	No
3262	Rubber Product Manufacturing	125	12	-	10	No
3271	Clay Product and Refractory Manufacturing	83	12	-	-	No
3272	Glass and Glass Product Manufacturing	153	41	-	-	Yes
3279	Other Nonmetallic Mineral Product Manufacturing	97	2	-	-	No
3311	Iron and Steel Mills and Ferroalloy Manufacturing	319	213	-	-	Yes
3312	Steel Product Manufacturing from Purchased Steel	330	235	-	-	Yes
3313	Alumina and Aluminum Production and Processing	307	23	-	-	No
3314	Nonferrous Metal (except Aluminum) Production and Processing	231	17	-	-	No
3315	Foundries	13	2	-	-	No
3321	Forging and Stamping	49	1	-	-	No
3322	Cutlery and Handtool Manufacturing	388	29	-	3	No
3323	Architectural and Structural Metals Manufacturing	97	2	-	-	No
3324	Boiler, Tank, and Shipping Container Manufacturing	186	5	-	-	No
3325	Hardware Manufacturing	83	5	-	-	No
3326	Spring and Wire Product Manufacturing	118	42	-	-	Yes

(continued on next page)

NAICS Code	Industry	Total Tariffs	AGOA Tariffs	Apparel Tariffs	Textiles Tariffs	Decision
3327	Machine Shops; Turned Product; and Screw, Nut, and Bolt Manufacturing	74	7	-	-	No
3328	Coating, Engraving, Heat Treating, and Allied Activities	128	4	-	3	No
3329	Other Fabricated Metal Product Manufacturing	404	38	-	3	No
3331	Agriculture, Construction, and Mining Machinery Manufacturing	202	2	-	-	No
3332	Industrial Machinery Manufacturing	546	20	-	14	No
3333	Commercial and Service Industry Machinery Manufacturing	439	2	-	3	No
3334	Ventilation, Heating, Air-Conditioning, and Commercial Refrigeration Equipment Manufacturing	496	20	-	15	No
3336	Engine, Turbine, and Power Transmission Equipment Manufacturing	114	3	-	-	No
3339	Other General Purpose Machinery Manufacturing	799	34	-	14	No
3342	Communications Equipment Manufacturing	129	4	-	-	No
3343	Audio and Video Equipment Manufacturing	180	36	-	-	No
3344	Semiconductor and Other Electronic Component Manufacturing	232	25	-	-	No
3345	Navigational, Measuring, Electromedical, and Control Instruments Manufacturing	575	139	-	4	No
3351	Electric Lighting Equipment Manufacturing	165	1	-	3	No
3352	Household Appliance Manufacturing	203	1	-	4	No
3353	Electrical Equipment Manufacturing	104	1	-	-	No
3359	Other Electrical Equipment and Component Manufacturing	104	3	-	-	No
3361	Motor Vehicle Manufacturing	27	20	-	-	Yes
3362	Motor Vehicle Body and Trailer Manufacturing	93	23	-	-	No
3363	Motor Vehicle Parts Manufacturing	266	7	5	9	No
3365	Railroad Rolling Stock Manufacturing	96	2	-	-	No
3366	Ship and Boat Building	360	23	-	14	No

(continued on next page)

NAICS Code	Industry	Total Tariffs	AGOA Tariffs	Apparel Tariffs	Textiles Tariffs	Decision
3369	Other Transportation Equipment Manufacturing	86	35	-	-	Yes
3371	Household and Institutional Furniture and Kitchen Cabinet Manufacturing	448	22	-	17	No
3372	Office Furniture (including Fixtures) Manufacturing	161	14	-	-	No
3379	Other Furniture Related Product Manufacturing	10	2	-	-	No
3391	Medical Equipment and Supplies Manufacturing	334	20	-	14	No
3399	Other Miscellaneous Manufacturing	771	59	11	64	No

Notes: This table provides a list of industries with at least one AGOA-eligible tariffline either in the general or the apparel or textile category. The "NAICS Code" column indicates each industry's classification code, while "Industry" provides the industry description. "Total Tariffs" shows the number of tariffs corresponding to the industry in 2000, and "AGOA Tariffs" indicates the count of AGOA-eligible tariffs within each industry. "Apparel Tariffs" and "Textiles Tariffs" specify the AGOA-eligible tariffs related to apparel and textiles, respectively. The "Decision" column denotes whether an industry qualifies as treated under AGOA, based on a threshold where at least 25% of tariff lines are AGOA-eligible, and categorizes the sector when applicable. Out of 93 total industries, 19 qualify under manufacturing, 8 under apparel or textiles, and 3 under agriculture; 63 industries do not meet the 25% threshold and are excluded from the analysis.

3

Beggar Thy Neighbor? Illicit Gold Trade and Conflict in the African Great Lakes Region

This chapter is co-authored with Jing-Rong Zeng

We examine the impact of Uganda's first gold refinery on conflict dynamics at artisanal gold mining sites in the Democratic Republic of Congo (DRC). Using a difference-in-differences approach and high-resolution data on mining activities and conflict events, we find that the refinery's establishment led to a significant increase in violence at neighboring artisanal mining sites within the DRC. To explore the mechanisms driving this increase, we constructed a unique dataset mapping the distribution of violent groups and incorporating road network data from both the DRC and Uganda. Our findings indicate that armed groups strategically targeted mining sites and key trade routes in response to the refinery's opening, underscoring the complex effects of mineral-related investments in fragile regions. This study highlights how weak state capacity and cross-border political-economic dynamics can amplify conflict and instability in resource-rich areas.

3.1 Introduction

Can policies exacerbate or attenuate violence in natural resources rich areas? Natural riches such as valuable minerals, oil or diamonds are frequently associated with heightened conflict in the regions where they are found, contributing to the well-established resource–conflict curse (Ross, 2004; Dube and Vargas, 2013; Berman et al., 2017)). However, the political economy mechanisms underpinning this relationship remain less well understood (Savoia and Sen, 2021). This paper proposes one mechanism: "beggar-thy-neighbor" dynamics, in which domestic policies in one country can generate conflict in neighboring resource-producing states.

Specifically, we examine whether the establishment of Uganda's first gold refinery in 2016, a facility that significantly boosted Uganda's gold exports and attracted international demand, affected conflict dynamics in the Democratic Republic of Congo (DRC), the primary source of Uganda's gold exports. Between 2015 and 2016, Uganda incentivized foreign investment in gold refining by offering tax exemptions on both raw gold imports and refined gold exports for at least ten years. This tax exemption policy lead to the establish of a gold refinery in Uganda in 2016. In this paper, we provide robust empirical evidence that Uganda's expanded refining capacity exacerbated conflict in artisanal and small-scale gold mining (ASGM) areas in eastern DRC, demonstrating a case of cross-border externalities in resource policy.

Our empirical strategy combines an original dataset from the International Peace Information Service (IPIS), which provides georeferenced data on artisanal and small-scale gold mining (ASGM) activity in eastern DRC, with conflict data from the Armed Conflict Location and Event Data (ACLED) project, which documents the location, type of conflict events, and actors involved. The units of analysis are 0.5×0.5-degree grid cells covering eastern DRC. We employ a difference-in-differences (DiD) design that exploits within-cell variation in conflict intensity before and after the establishment of the refinery in 2016. Treatment cells are defined as those containing at least one ASGM site prior to the refinery's establishment. To further capture spatial variation in exposure to Uganda's gold refinery, we construct a road network using data from OpenStreetMap and compute the travel distance from each cell to the refinery. We hypothesize that cells located closer to the refinery are more likely to lie along gold smuggling routes and, consequently, be more exposed to conflict. To strengthen our identification, we control for time-varying macroeconomic shocks, including international gold price fluctuations, as well as weather-related shocks such as rainfall and temperature.

Our findings reveal a significant increase in conflict intensity around ASGM sites near the Uganda-DRC border after the refinery's establishment. Specifically, we observe a 12.8% increase in the likelihood of violent battles, a 37.5% increase in conflict events, and a 45.3% rise in battle-related fatalities in affected regions post-2016. These results remain robust when controlling for global gold price fluctuations and when accounting for spatial and temporal clustering of standard errors.

Our analysis suggests that Uganda's formalization of gold refining and export has intensified violence by increasing the incentives for appropriation by armed groups. In a context characterized by weak property rights and limited state enforcement, looting and controlling existing production sites can be more attractive than engaging in labor-intensive extraction. This tendency is particularly pronounced in fragile cross-border regions, where porous borders further reduce the costs of violent appropriation and complicate enforcement efforts.

This paper makes several contributions. First, it extends the literature on natural resource extraction and conflict by shifting the focus from temporary price shocks to structural, trade-based shocks (Berman et al., 2017; Dube and Vargas, 2013; Sánchez De La Sierra, 2020). While existing studies emphasize how fluctuations in commodity prices can increase the incentives and feasibility of violence, we show that the opening of a gold refinery in Uganda contributed to intensified conflict in artisanal and small-scale mining (ASGM) sites in eastern DRC by increasing illicit gold trade. We interpret the establishment of the refinery as a permanent structural shock, distinct from temporary price shocks, with more persistent effects that can transform the political economy of the region.

Second, this study also introduces a novel application of the "beggar-thy-neighbor" concept to the political economy of conflict. We highlight how Uganda's pursuit of economic gains through gold refining can have destabilizing effects on its neighbors, suggesting that regional coordination is essential to address the cross-border dimensions of the resource curse.

Finally, we contribute to research on illicit economies and transnational crime, by documenting how cross-border smuggling can undermine international transparency regulations such as Section 1502 of the Dodd-Frank Act (Dell, 2015; Justino et al., 2024; Zabyelina and Heins, 2020). Our findings help explain why well-intentioned reforms may fall short when enforcement is weak and minerals with high value-to-volume ratio can be laundered through neighboring countries with refining capacity.

The remainder of the paper is organized as follows. Section 3.2 provide context of the ASGM sites in Eastern DRC and the dynamics of the region. Section 3.3 develops the conceptual framework of the paper. Section 3.4 discusses the details of our data. Section 3.5 displays the empirical strategy, and Section 3.6 presents the baseline results. Section 3.7 presents several robuteness and Section 3.8 concludes.

3.2 Background

3.2.1 Artisanal Gold Mining in Eastern DRC

Artisanal and small-scale mining (ASM) is defined as "formal or informal mining operations with predominantly simplified forms of exploration, extraction, processing, and transportation (Gillard, 2015)." This type of mining typically involves low capital investment and relies heavily on manual labor, with operations that may be either legally recognized or informal in nature (Girard et al., 2022). Around 60 million people across Sub-Saharan Africa rely on this labor-intensive method for extracting minerals (Hilson et al., 2017).

In the DRC, 64% of the ASM sites primarily extract gold, with tin coming in second place at 22% and coltan in third at 6%.[1] The Ministry of Mines of DRC permits artisanal mining within specific zones designated as artisanal exploitation areas, although this regulation is often poorly enforced.

Artisanal and small-scale gold mining (ASGM) in the Eastern DRC provides vital income sources in regions with scarce economic opportunities. In contrast, large-scale industrial mining struggles to integrate with the local economy (Bazillier and Girard, 2020) and typically offers fewer employment opportunities (Christiaensen and Martin, 2018).

[1] Tin (cassiterite), tungsten (wolframite), and tantalum (coltan) are the minerals collectively referred to as the 3Ts, also often linked to conflicts. The summary of primary minerals extracted in Eastern DRC is based on the authors' calculations and the map surveyed by (Müller, 2015).

A 2019 survey estimated that at least 274,000 individuals were employed in gold mining in Eastern DRC.[2] Although artisanal mining provides a relatively high income compared to other local jobs, miners often face dangerous working conditions, with little regard for safety, health, or environmental protections.

Despite an estimated annual gold production of 13 tonnes , official records from 2017 report only 230 kilograms of artisanal gold exports, highlighting the significant extent to which ASGM production in the DRC bypasses formal state oversight and trade channels (Ober, 2018).

3.2.2 Cross-Border Trade and Gold Refining in the Great Lakes Region

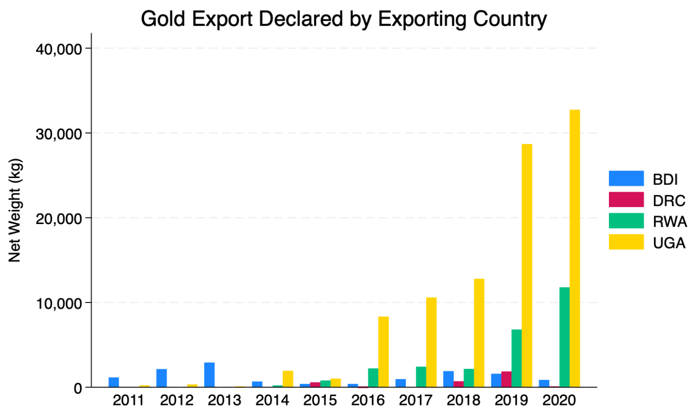


Figure 3.1. Declared Gold Exports by Country in the Great Lakes Region (2011–2020)

Notes: This figure shows the quantity of gold exports declared by each country: Burundi (BDI), Democratic Republic of Congo (DRC), Rwanda (RWA), and Uganda (UGA) between 2011 and 2020. Data before 2015 contain missing values for several countries. Export quantities are calculated by summing reported values across all trading partners, excluding the aggregate category "WORLD."
Sources: UN Comtrade Database (commodity code 7108)

[2]Authors' calculation based on (Muller, 2020).

Eastern DRC is economically far more integrated with its eastern neighbors—particularly Uganda and Rwanda—than with the distant national capital, Kinshasa (Titeca and Salvaggio, 2025). This cross-border integration is reflected in both trade flows and infrastructure: most imports into Eastern DRC arrive via Uganda and Rwanda, and key commodities such as gold and coffee are also primarily exported through these countries rather than to the rest of the DRC.

Figure 3.1 shows declared gold exports from Great Lakes countries between 2011 and 2020. Uganda's reported gold exports increased sharply from 1,026 kg in 2015 to 8,337 kg in 2016. Given Uganda's limited domestic gold production, such a surge is unlikely to reflect domestic output. Instead, this increase coincides with reports that "gold from the Democratic Republic of the Congo (DRC) is mainly smuggled towards East Africa, mainly to Uganda, Rwanda, and Burundi.[3]

A key factor in this export boom was the opening of Uganda's first major refining facility, the African Gold Refinery (AGR), in 2016. As noted by the UN Group of Experts, "the revival of official gold exports from Uganda coincides with the 2016 opening of the country's first gold refinery, African Gold Refinery (AGR)".[4] A gold refinery processes unrefined or semi-refined gold sourced from artisanal or small-scale miners and purifying it into high-grade bullion suitable for export or resale. The refinery is strategically located near Entebbe International Airport, just outside the capital, Kampala. Given the high value-to-volume ratio of refined gold, most exports are transported by air.

Uganda pursued the gold refining industry to "stimulate investment in the mineral sector by promoting private participation", and "to ensure that mineral wealth supports national economic and social development". In addition, refining gold adds value to exports, as "refined gold products naturally fetch a higher value when exported than the unprocessed gold.[5]

[3] "From these transit countries, gold is mainly exported or smuggled towards the United Arab Emirates (UAE), and to Europe and Asia as final destinations (*Illegal Gold Mining in Central Africa*, n.d.)."

[4] UN Security Council Report S/2017/672/Rev.1

[5] The United Nations Economic Commission for Africa provides a summary of Uganda's artisanal mining strategy and policy, available at https://www.uneca.org/."

3.2.3 Armed Groups and Resource Exploitation

Both state-affiliated and non-state armed actors play a critical role in the artisanal mining sector in Eastern DRC. According to the IPIS survey, 61% of artisanal mining sites report interference by armed groups, and 37% indicate a permanent presence of these actors (Matthysen et al., 2023). State actors, such as the Armed Forces of the Democratic Republic of the Congo (FARDC), “were reportedly illegally owning and taxing gold mines, employing between 15,000 and 20,000 diggers, in Mongbwalu", a town in Ituri. Non-state armed groups also “draw their resources from the gold mining sector, through attacks on facilities, control of sites and gold flows, and illegal taxation."[6]

Artisanal miners capture only a small portion of the gold’s full value. A substantial part of the profits benefits armed groups, local elites, and transnational criminal networks that exploit the informal nature of the sector.[7] Evidence of exploitation is reflected in the local gold buying prices, too. While international and local gold prices tend to follow similar trends (Sánchez De La Sierra, 2020), “conflict actors are reported to offer 60% or less of the London Bullion Market Association value in sites and territories they control", compared to 75-95% elsewhere in Africa (Hunter, 2019).

3.3 Conceptual Framework

While the establishment of a gold refinery in Uganda marks a significant boom in the regional gold economy, its implications for conflict dynamics in Eastern DRC are theoretically ambiguous. The literature on insurgency participation highlights several competing mechanisms. On one hand, a positive income shock in a labor-intensive sector may reduce violence by raising the opportunity cost of participation in conflict (*opportunity cost effect*). On the other hand, a shock to natural resource income can increase violence by raising the value of contestable rents, thereby incentivizing appropriation by armed actors (*rapacity effect*). (Dube and Vargas, 2013) empirically distinguish these channels using price shocks to coffee (a labor-intensive crop) and oil (a capital-intensive resource) in Colombia.

[6] (*Illegal Gold Mining in Central Africa*, n.d.)

[7] INTERPOL estimates that illicit profits from the gold trade alone can finance up to 8,000 combatants, supporting between 25 and 49 armed groups in the region (*Illegal Gold Mining in Central Africa*, n.d.).

3. Beggar Thy Neighbor? Illicit Gold Trade and Conflict in the African Great Lakes Region

Artisanal gold mining in Eastern DRC is labor-intensive, and a positive income shock in principle could increase miners' earnings and reduce incentives to engage in violence. However, as (Dal Bó and Dal Bó, 2011) show in a general equilibrium model, even in labor-intensive sectors, weak property rights and limited enforcement can lead to higher conflict when appropriation becomes more attractive than production. In such settings, increases in resource value can shift incentives toward appropriation, rather than away from violence.

Whether the gold refinery ultimately increases or reduces violence depends on the incentives it generates, particularly for actors engaged in appropriation. Prior to the refinery's opening, cross-border flows of artisanal gold already existed, but much of this trade was informal and unrecorded. The refinery made it possible to legalize previously illicit gold exports by offering a route to international markets without full traceability of origin. In doing so, the refinery effectively improved market access for gold sourced from conflict-affected zones.[8] Moreover, refining increases the value of gold by transforming it into a higher-grade export product. This creates an additional, durable income shock over and above fluctuations in the international gold price.

These dynamics generate several testable implications. First, the opening of a gold refinery that does not verify the origin of gold increases the profitability of appropriation and may therefore lead to higher levels of violent conflict in gold-producing areas. Second, the impact of the refinery should be spatially heterogeneous: the closer a mining area is to the refinery, the greater the gain in market access and, consequently, the stronger the incentive for contestation. This implies that the refinery's effect on violence should attenuate with distance from the refinery. Third, we expect the refinery to have a stronger impact on the *intensity* of violence, measured by the number of conflict events or fatalities, than on the *likelihood* of conflict onset. This is because the refinery does not create new conflicts per se, but increases the *feasibility* of violence in areas where conflict already exists, by making gold extraction and smuggling more profitable (Berman et al., 2017). Therefore, the value-added and (illicit) trade facilitation effects of the refinery are more likely to escalate ongoing violence in Eastern DRC than to trigger new outbreaks.

In addition to the distinction between likelihood and intensity of conflict, we

[8] Although the UN Group of Experts "believes that African Gold Refinery (AGR) could contribute to a cleaner gold trade in Uganda" by formalizing gold exports, it also noted that, as of 2017, AGR "does not have the capacity to map the actual physical flow of each and every gram of gold." See S/2017/672/Rev.1.

also differentiate between existing gold mines and new mines opening after the refinery shock to shed lights on the opportunity cost channel. This distinction is useful for thinking about the opportunity cost channel, even if it is harder to empirically isolate. Artisanal gold mining is labor-intensive and, although low in capital intensity, often requires fixed investments and some degree of local organization including pit managers or informal financing arrangements. In such settings, persistent violence can undermine production, suggesting that new mining areas may be more sensitive to opportunity cost considerations and the opportunity cost channel may dominate the rapacity effect. We elaborate on this mechanism further in Appendix Section 3.B.3.

3.4 Data

We divide the Democratic Republic of Congo (DRC) into grid cells measuring $0.5° \times 0.5°$ in latitude and longitude, and construct a grid-year panel dataset covering the period from 2011 to 2020. At the equator, each grid cell corresponds to approximately 55 km $\times$ 55 km. Grid cells that intersect national borders are clipped to align precisely with the boundaries, resulting in smaller, irregularly shaped units in those areas. The final dataset includes a total of 850 grid cells across 10 years.

We adopt a grid structure rather than subnational administrative boundaries to mitigate potential endogeneity concerns related to both conflict dynamics and the definition of political or administrative units (Besley and Reynal-Querol, 2014; Berman et al., 2017; Michalopoulos and Papaioannou, 2013).

3.4.1 Artisanal Gold Mining Location

We identify the geographic locations of artisanal mining sites using data from the International Peace Information Service's (IPIS) project, "Mapping Artisanal and Small-Scale Mining" in Eastern DRC. Since 2009, IPIS has conducted field visits and surveys using GPS devices to document the locations of artisanal mining sites. The dataset provides broad coverage across the eastern provinces, including Ituri, North Kivu, and South Kivu.[9]

[9] For more details on the IPIS survey methodology, see Appendix Section 3.A.1. While other studies, such as (Girard et al., 2022), use geological gold suitability to proxy for potential artisanal gold mining activity, this approach is less precise in the DRC context.

Our analysis focuses exclusively on the geographic presence of artisanal gold mining activity, abstracting from the timing of survey visits. However, to define treatment, we restrict attention to grid cells in which artisanal gold mining sites were identified on or before 2016 (the year when the Ugandan gold boom began). A total of 96 out of 850 grid cells meet this criterion and are classified as treated.[10]

3.4.2 Travel Distance

We obtain road network data from OpenStreetMap[11] and calculate the travel distance from each grid cell centroid to the gold refinery located in Entebbe, Uganda. Given the large geographic scale of the DRC and significant variation in road quality across regions, we restrict our analysis to major roads, including primary, secondary, and roads linking these primary and secondary network. These road types generally correspond to national highways or provincial motorways and are less likely to be impassable due to poor maintenance or seasonal degradation (Lebrand et al., 2025). Using network analysis, we compute the shortest travel distance along the connected road network from each grid cell centroid to the refinery, capturing spatial variation in accessibility across the study area.[12]

3.4.3 Outcome Data

We use the widely-adopted Armed Conflict Location & Event Data (ACLED) to identify geo-coded violent events and their various types (Raleigh et al., 2023). ACLED classifies conflict events into several major categories: battles,

---

Many geologically suitable areas are inaccessible to mining due to factors such as national parks, government-imposed restrictions, or dense rain forests. Therefore, we rely on observed mining site data from IPIS and do not use geological suitability in our definition of treatment.

[10] As a robustness check, we compare results based on this treatment definition with an alternative, broader definition that includes all grid cells where artisanal gold mining activity was ever recorded between 2009 and 2023. There are 22 grid cells where gold mines were identified exclusively after 2016. Additional discussion and results are provided in Appendix Section 3.B.3.

[11] Data accessed in September 2024 from OpenStreetMap GeoFabrik (*Geofabrik Download Server*, n.d.; OpenStreetMap contributors, n.d.). For details on data processing and implementation, see Appendix Section 3.A.2.

[12] A map illustrating travel distances and the underlying road network is provided in Figure 3.A.1, located in Appendix Section 3.A.2.

explosions/remote violence, protests, riots, strategic developments, and violence against civilians. During the study period, 34.7% of the recorded events were battles, 33.9% were violence against civilians, followed by 12.5% protests, 10.2% strategic developments, and 8% riots. Remote violence, such as the use of improvised explosive devices, is relatively rare in the DRC. We focus primarily on battle-related events and violence against civilians, which together account for more than half of all recorded violent events during the study period.

3.4.4 Additional Data

In addition to data on artisanal mining and conflict events, we incorporate several complementary data sources that provide key control variables for our analysis. Given the well-established relationship between climate shocks and conflict (Burke et al., 2015), we include grid-level measures of rainfall and temperature anomalies. These climate shocks also influence livelihood decisions: artisanal miners may shift from agricultural activities to mining during periods of drought or extreme temperature, using mining as an alternative source of income (Banchirigah and Hilson, 2010).[13] To account for these time-varying effects, we include grid-cell-level climatic shocks in our regression framework. Climatic shocks are defined as standard deviations from the yearly average in precipitation and temperature, adjusted for seasonal variations.[14]

Table 3.1 presents summary statistics for the key variables, and Figure 3.2 plots the geographic distribution of conflict events before and after the opening of the gold refinery. Artisanal gold mining grid cells are concentrated in Eastern DRC, with an average travel distance of 973.5 kilometers to the refinery. Conflict events are similarly clustered in Eastern DRC, particularly along the borders with Uganda and Rwanda. ACLED records a total of 4,580 conflict events between 2011 and 2015, and 9,959 events between 2016 and 2020.

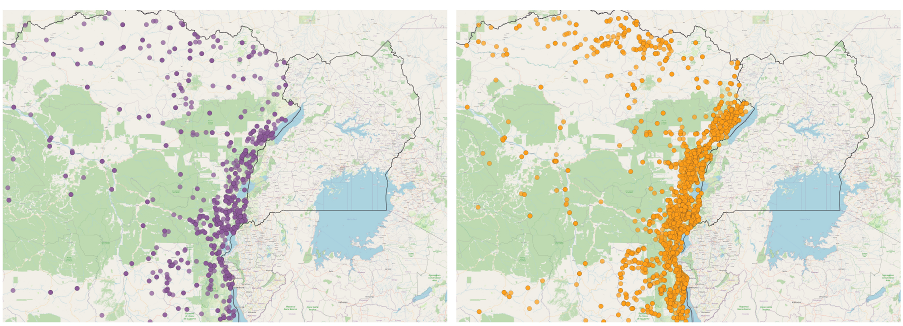
[13] To proxy for artisanal gold production, we use high-resolution deforestation data from (Hansen et al., 2013), which captures land-use changes often associated with ASGM activity (Girard et al., 2022). We present descriptive trends on deforestation in Appendix Section 3.B.2.

[14] We follow the approach of (Desmet et al., 2024) in incorporating time-varying cell-level measures.

Table 3.1. Summary Statistics

Variable	N	Mean	SD	p25	Median	p75
Count of Conflict Events	8500	1.71	11.08	0	0	0
Pr(Events > 0)	8500	0.18	0.38	0	0	0
Count of Battles	8500	0.6	4.38	0	0	0
Pr(Battles > 0)	8500	0.10	0.30	0	0	0
Count of VAC	8500	0.58	4.16	0	0	0
Pr(VAC > 0)	8500	0.11	0.31	0	0	0
Travel Distance to Refinery	8500	1702	585	1283	1709	2050
Travel Distance to Refinery (ASGM)	960	973	255	751	978	1096
Rainfall (total annual)	8500	1375	621	1208	1349	1507
Temperature (average annual)	8500	24.54	1.78	23.67	25.04	25.66

Notes: This table presents summary statistics for 850 grid cells from 2011 to 2020. Each cell has a resolution of $0.5° \times 0.5°$ (approximately 55 km by 55 km at the equator). Travel Distance to Refinery is measured in kilometers. Cells located along country borders are split according to national boundaries to reflect realistic territorial divisions.



(a) Pre-Refinery Period (2011-2015) **(b)** Post-Refinery Period (2016-2020)

Figure 3.2. Locations of ACLED Events (All Types)

Notes: The map is zoomed in on Eastern DRC. Each dot represents a conflict event, with transparency set to 70%. Areas with higher concentrations of events appear darker due to overlapping points. There are 4,580 events recorded by ACLED during 2011 and 2015 and 9,959 events recorded during 2016 and 2020.

3.5 Empirical Strategy

3.5.1 Main Specification

$$\begin{aligned} Y_{it} = \alpha + \beta_1 \mathrm{ASGM}_i \times \mathrm{PostRefinery}_t + \beta_2 \mathrm{ASGM}_i \times \mathrm{Dist}_i \times \mathrm{PostRefinery}_t \\ + \ \mathbf{X}_{it}\Gamma + \gamma_i + \gamma_{rt} + \varepsilon_{it}, \end{aligned} \tag{3.1}$$

Subscripts i and t index cells and time periods. Y_{it} denotes conflict-related outcomes in grid cell i and year t. ASGMi is a binary indicator equal to one for cells with artisanal gold mines ever recorded in or before 2016, and zero otherwise. Post_t is an indicator equal to one for the post-treatment period (2016–2020), following the gold boom induced by the expansion of gold refining capacity, and zero otherwise. Dist_i is a continuous variable measuring the travel distance from cell i to the gold refinery in Uganda. $\mathbf{X}_{it}$ includes time-varying climatic controls such as temperature and rainfall shocks. γ_i denotes grid-cell fixed effects, and γ_{rt} captures region-year fixed effects, where regions are defined by provincial borders as of 2015, allowing for differential conflict dynamics across regions in the Democratic Republic of Congo (DRC), a vast country with considerable heterogeneity from Kinshasa to the East. These fixed effects control for time-invariant unobserved heterogeneity across cells, and account for region shocks that affect all the cells within a region. ε_{it} is the error term.

We have two main variables of interest. The first is the interaction between the existence of artisanal and small-scale gold mining and the post-refinery indicator. The coefficient on this interaction, β_1, captures the difference-in-differences estimate of the average effect of the gold refinery on conflict outcomes in cells that had an ASGM site prior to the refinery's establishment The second variable of interest is the interaction of the first one with the distance from the cell to the refinery. Its coefficient, β_2, captures how the effect of the refinery varies with distance among ASGM cells. We expect β_1 to be positive, as cells with ASGM sites prior to the refinery's establishment should experience a greater increase in conflict. In contrast, we expect β_2 to be negative, since the effect of the refinery on conflict should diminish as the distance from the refinery increases.

The identifying assumption required for unbiased estimation of β_1 and β_2 is a generalized parallel trends assumption. Specifically, in the absence of the refinery, (i) the trends in conflict outcomes for ASGM and non-ASGM cells would have evolved similarly over time, and (ii) among ASGM cells, conflict trends would not have varied systematically with distance to the refinery.

3.5.2 Event-Study specification

We examine for the parallel pretreatments trends assumption using an event study approach. To incorporate the effect of distance to the refinery, we divide the sample by the median, separating those cells which are closer to the refinery and those cells which are further of it. Following Jacobson et al. (1993), we estimate the following equation:

$$Y_{it} = \alpha + \sum_{\mathrm{k} \geq -4, \mathrm{k} \neq -1}^{k=4} \left[\delta_k^{\text{Near}} \cdot D_{i,t-k} \cdot \text{Near}_i + \delta_k^{\text{Far}} \cdot D_{i,t-k} \cdot \text{Far}_i \right] + \mathbf{X}_{it}\Gamma + \gamma_i + \gamma_{rt} + \varepsilon_{it}, \tag{3.2}$$

where Y_{it} denotes the conflict outcome in cell i at time t. The dummy variable $D_{i,t-k}$ is a dummy indicating event time k, i.e., k years relative to the establishment of the refinery, 2016. Note that the dummy for $k = -1$ is omitted, so all estimated effects are relative to the year immediately preceding the refinery's establishment. The variables Near_i and Far_i are indicators for whether a cell lies below or above the median distance to the refinery, respectively.

Our parameters of interest are δ_k^{Far} and δ_k^{Near}, which estimate the effect of the refinery's establishment k years after the event for cells near to and far from the refinery, respectively. We also include leads ($k < 0$) to serve as placebo tests for the parallel trends assumption. The vector X_{it} includes time-varying controls for rainfall and tempreature shocks. Finally, γ_i is a cell fixed effect, while γ_{rt} is a region-year fixed effect.

3.6 Results

This section presents the main results of the paper, including the dynamic treatment effects from the event study analysis.

3.6.1 Baseline Results

We examine the effect of the establishment of the refinery on conflict. We focus on three main conflict outcomes: the probability of a violent event, the total number of violent events, and the total number of fatalities. As discussed above, we define treated cells as those with at least one ASGM site opened before the establishment of the refinery. Since new ASGM mines opened after the refinery's establishment, potentially contaminating the control group, we report results using two different samples for each outcome: one that includes cells with no ASGM before 2016 but with at least one mine opened after 2016, and another that excludes these cells.

The results are presented in Table 3.3. Columns (1) and (2) report the effect of the refinery's establishment on the probability of a conflict event; Columns (3) and (4) report effects on the total number of violent events; and Columns (5) and (6) focus on the number of fatalities. Standard errors are reported below the estimated coefficients. Across all specifications, we find that the establishment of the refinery has a statistically significant and positive effect on conflict in grid cells that hosted at least one ASGM site prior to 2016. We interpret this as evidence that the refinery introduced a persistent income shock, leading to increased violence along the intensive margin of conflict, namely, more events and higher fatalities.

Furthermore, we observe that this effect is attenuated by travel distance to the refinery. Since distance is measured in units of 100 kilometers, the estimated coefficients imply that moving from the 25th percentile (751 km) to the 75th percentile (1,097 km) – an increase of approximately 346.5 kilometers – is associated with a decrease of about 0.27 in the IHS-transformed number of events and a decrease of approximately 0.36 in the IHS-transformed number of battle fatalities. This attenuation pattern suggests that the refinery's impact on conflict intensity weakens considerably with increasing distance.

Table 3.4 follow the same structure than Table 3.3 but focus on violence against civilians. The reason why we specifically look at this type of violence is because ASGM is a labor-intense sector, and violent groups use civilians to extract the gold from the mining. As reported, we observe a positive and significant increase in violence against civilians.

Table 3.3. Effect of Opening Gold Refinery on Battle

Type of Conflict	Battle					
Dependent Var.	Pr(event > 0)		IHS Events Counts		IHS Fatalities	
	(1)	(2)	(3)	(4)	(5)	(6)
ASGM×PostRefinery	0.264*	0.270*	0.881**	0.898**	1.308***	1.322***
	(0.107)	(0.106)	(0.282)	(0.281)	(0.388)	(0.387)
ASGM×Dist×PostRefinery	-0.017	-0.017	-0.071**	-0.071**	-0.103**	-0.103**
	(0.011)	(0.011)	(0.026)	(0.025)	(0.034)	(0.034)
Mean of Dep. Var.	0.098	0.096	0.173	0.171	0.195	0.193
Cell FE	Yes	Yes	Yes	Yes	Yes	Yes
Region-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Cells with New Mines	Yes	No	Yes	No	Yes	No
Observations	8500	8280	8500	8280	8500	8280
Adjusted R^2	0.449	0.456	0.682	0.690	0.562	0.570

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Robust standard errors clustered at the grid-cell level are reported in parentheses. Distance is expressed in units of 100 kilometers. Columns (1), (3), and (5) are estimated using equation (3.1) with the full sample, while columns (2), (4), and (6) exclude grid cells where ASGM sites were recorded only after (and excluding) 2016. The dependent variable in the linear probability model, Pr(event > 0), equals 1 if any conflict event (battle) occurred in a given year, and 0 otherwise. IHS refers to the inverse hyperbolic sine transformation.

Table 3.4. Effect of Opening Gold Refinery on Violence Against Civilians

Type of Conflict	Violence Against Civilians					
Dependent Var.	Pr(event > 0)		IHS Events Counts		IHS Fatalities	
	(1)	(2)	(3)	(4)	(5)	(6)
ASGM×PostRefinery	0.244*	0.245*	0.846**	0.845**	1.066**	1.079**
	(0.120)	(0.119)	(0.300)	(0.301)	(0.362)	(0.362)
ASGM×Dist×PostRefinery	-0.020	-0.020	-0.078**	-0.078**	-0.084**	-0.084**
	(0.011)	(0.011)	(0.027)	(0.027)	(0.031)	(0.031)
Mean of Dep. Var.	0.106	0.104	0.181	0.179	0.145	0.145
Cell FE	Yes	Yes	Yes	Yes	Yes	Yes
Region-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Cells with New Mines	Yes	No	Yes	No	Yes	No
Observations	8500	8280	8500	8280	8500	8280
Adjusted R^2	0.478	0.484	0.670	0.677	0.522	0.527

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Robust standard errors clustered at the grid-cell level are reported in parentheses. Distance is expressed in units of 100 kilometers. Columns (1), (3), and (5) are estimated using equation (3.1) with the full sample, while columns (2), (4), and (6) exclude grid cells where ASGM sites were recorded only after (and excluding) 2016. The dependent variable in the linear probability model, Pr(event > 0), equals 1 if any conflict event (violence against civilians) occurred in a given year, and 0 otherwise. IHS refers to the inverse hyperbolic sine transformation applied to continuous, non-negative outcomes to approximate log transformation while preserving zeros.

3.6.2 Dynamics and Heterogeneity

In this section, we assess the validity of the parallel trends assumption underlying our difference-in-differences identification strategy, we explore the dynamic effects of the refinery's establishment using an event study framework. This approach allows us to trace the evolution of conflict outcomes both before and after the treatment and to test for the presence of any pre-treatment trends.

Table 3.5 reports the regression results from Equation 3.2, using the probability of violence and the total number of fatalities as conflict outcomes. The sample is divided between cells that are near to and far from the refinery, based on if an ASGM grid cell falls below or above the median distance. Figures 3.3 and 3.4 visualize the dynamic treatment effects for both outcomes, along with their 95% confidence intervals. Each dot represents the estimated coefficient of the treatment dummy for a given number of years before or after the establishment of the gold refinery.

Across all four columns, the estimated coefficients on the leads of treatment are statistically indistinguishable from zero, indicating no significant pre-treatment effects. Visual inspection of Figures 3.3 and 3.4 further supports the absence of pre-trends. Together, this evidence suggests that pre-treatment trends in conflict are similar between treated and untreated cells, lending credibility to the parallel trends assumption required for unbiased estimation in the difference-in-differences framework. Similar results for violence against civilians are presented in Appendix Section 3.B.1.

Another important point to highlight is the divergent post-treatment trends between areas located near the refinery and those farther away. We observe that the increase in conflict in ASGM regions is more pronounced in cells closer to the refinery. This effect is particularly evident in Figure 3.4, which focuses on the intensity of conflict, as opposed to Figure 3.3, which examines the probability of violence. These findings are consistent with our main hypothesis and support the results reported in Table 3.3, which show that distance from the refinery has a mitigating effect on violence.

Table 3.5. Tests for the Parallel Trends Assumption (Outcome: Battle)

Event Time	(1) Pr(Battle>0) Far	(2) Pr(Battle>0) Near	(3) IHS(Battle) Far	(4) IHS(Battle) Near
4 years before	-.033 (.047)	.042 (.053)	-.122 (.08)	.152 (.168)
3 years before	.055 (.06)	-.003 (.056)	.089 (.137)	.299 (.178)*
2 years before	.138 (.063)*	-.041 (.054)	-.047 (.126)	.174 (.166)
Year Refinery	.038 (.056)	.084 (.06)	-.078 (.098)	.349 (.134)**
1 year after	.058 (.047)	.031 (.055)	.22 (.161)	.304 (.169)*
2 years after	-.018 (.051)	.121 (.058)*	.024 (.189)	.55 (.224)**
3 years after	.035 (.056)	.138 (.073)*	-.131 (.11)	.643 (.244)**
4 years after	.023 (.067)	.083 (.077)	.176 (.157)	1.431 (.274)***

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Tabulated results correspond to Figures 3.3 and 3.4. Robust standard errors clustered at the grid-cell level are in parentheses. Dependent variables include a binary indicator for any conflict event or the IHS-transformed number of events or fatalities. Models are estimated jointly by equation (3.2). "Near" and "Far" refer to grid cells below and above the median distance to the border.

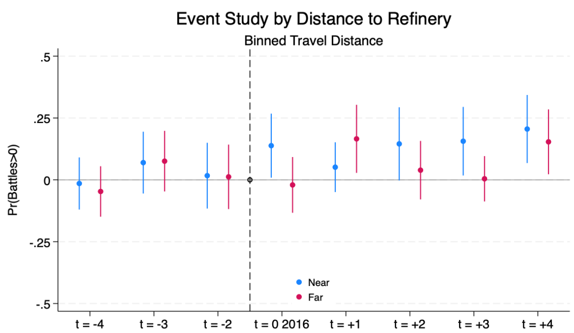


Figure 3.3. Event Study of Battle Likelihood by Distance to Refinery

Notes: This figure plots the coefficients and associated 95% confidence intervals estimated using equation (3.2). The dependent variable is binary and equals one if at least one battle event occurred in the grid in a given year, and zero otherwise. Distance to the gold refinery is divided into "near" and "far" based on whether an ASGM grid cell falls below or above the median distance. We examine heterogeneity by distance to explore how proximity to the refinery may influence smuggling incentives and associated conflict dynamics following its opening.

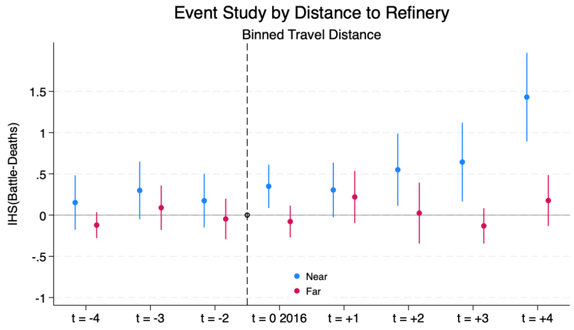


Figure 3.4. Event Study of Battle-related Deaths by Distance to Refinery

Notes: This figure plots the coefficients and associated 95% confidence intervals estimated using equation (3.2). The dependent variable is the inverse hyperbolic sine (IHS) transformation of deaths resulting from battles. Distance to the gold refinery is divided into "near" and "far" based on whether an ASGM grid cell falls below or above the median distance. We examine heterogeneity by distance to explore how proximity to the refinery may influence smuggling incentives and associated conflict dynamics following its opening.

3.7 Robustness

3.7.1 Spatial and Temporal Correlation

Using grid cells as the spatial unit of analysis helps mitigate concerns about endogenously determined administrative boundaries that may be correlated with conflict outcomes. However, a positive income shock such as the one induced by the opening of the gold refinery can affect neighboring units. Moreover, the refinery represents a persistent shock to artisanal gold export opportunities, and conflict outcomes are likely to exhibit temporal autocorrelation. To address these concerns, we implement Conley-Hsiang standard errors (Conley, 1999; Hsiang, 2010), which allow for correlation to decay smoothly over both space and time.

Following prior studies on conflict in Sub-Saharan Africa (Stoop et al., 2018; Berman et al., 2017), and given the large geographic size of the DRC, we apply a spatial cut-off of 500 kilometers and allow for temporal correlation across 10 periods, corresponding to the full duration of our panel from 2011 to 2020. Table 3.6 presents the results on battle outcomes estimated using equation (3.1), with Conley-Hsiang standard errors. The estimated effect of the gold refinery opening on battle intensity, as well as the heterogeneous effect by travel distance, remains robust and statistically significant, although standard errors are slightly larger.

Table 3.6. Effect of Opening Gold Refinery on Battle

Type of Conflict	Battle					
Dependent Var.	Pr(event > 0)		IHS Events Counts		IHS Fatalities	
	(1)	(2)	(3)	(4)	(5)	(6)
ASGM×PostRefinery	0.264*	0.270**	0.881**	0.898**	1.308**	1.322**
	(0.106)	(0.101)	(0.293)	(0.285)	(0.459)	(0.451)
ASGM×Dist×PostRefinery	-0.017	-0.017	-0.071*	-0.071**	-0.103**	-0.103**
	(0.011)	(0.011)	(0.028)	(0.027)	(0.040)	(0.040)
Cell FE	Yes	Yes	Yes	Yes	Yes	Yes
Region-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Cells with New Mines	Yes	No	Yes	No	Yes	No
Observations	8500	8280	8500	8280	8500	8280

3.7.2 International Price Shocks

Our conceptual framework emphasizes that the opening of the gold refinery introduces a persistent income shock by adding value to gold exports. This increases the value of contestable resources beyond what is captured by fluctuations in international gold prices, thereby amplifying the rapacity effect relative to the opportunity cost effect in artisanal gold mining areas. Figure 3.5 compares Uganda's declared gold export prices with international gold prices reported by the World Bank. The figure shows that, beginning in 2015, Uganda's export prices increasingly track international prices in both levels and trends while its gold export prices between 2011 to 2013 were much lower significantly lower than international benchmarks.[15]

In this robustness check, we examine whether our main results remain robust when controlling for fluctuations in international gold prices. Building

[15] One may suspect that Ugandan traders had incentives to under-report the quantity of gold exported prior to the refinery's opening to avoid regulatory oversight. However, under-reporting quantities does not necessarily lead to a downward bias in export prices. We interpret the observed increase in export prices as reflecting improvements in the quality and purity of gold being exported.

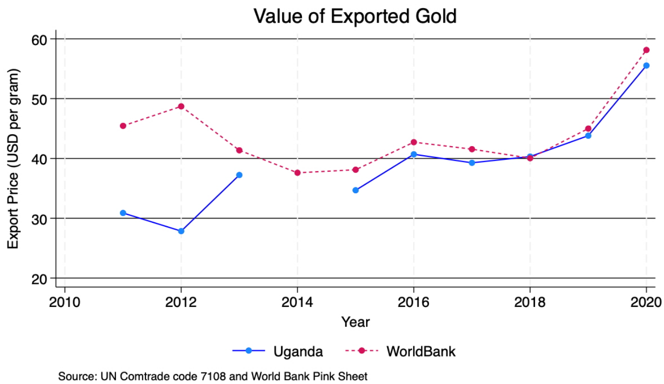


Figure 3.5. Trends in International Gold Prices and Ugandan Export Prices

Notes: International gold price data are sourced from the World Bank's Pink Sheets and are expressed in real U.S. dollars. Prices are converted from U.S. dollars per ounce to dollars per gram. Uganda's gold export price is approximated by dividing Uganda's self-reported export values (in dollars) by the corresponding reported export quantities (in grams). Export data for Uganda are missing for the year 2014.
Source: UN Comtrade Database and World Bank Pink Sheet

on equation (3.1), we include the logarithm of international gold prices interacted with artisanal gold mining cells to test whether global price variation can account for the observed post-refinery effects. This allows us to assess whether the refinery's impact on conflict outcomes in ASGM areas is fully explained by international price movements, or whether it represents an independent and persistent export shock.

Formally, we estimate the regression specified in equation (3.3). Drawing on the literature on international commodity price shocks and conflict over natural resources (Berman et al., 2017), we expect β_3, the coefficient on the interaction between international log gold price and ASGM areas, to be positive. Our primary interest remains in whether β_1, which captures the post-refinery effect in artisanal gold mining areas, remains positive and statistically significant after accounting for global price fluctuations.

$$\begin{aligned} Y_{it} &= \alpha + \beta_1 \mathrm{ASGM}_i \times \mathrm{Post}_t + \beta_2 \mathrm{ASGM}_i \times \mathrm{Dist}_i \times \mathrm{Post}_t + \beta_3 \ln(\mathrm{Price}_t) \times \mathrm{ASGM}_i \\ &\quad + \mathbf{X}_{it}\Gamma + \mathrm{FE}_i + \mathrm{FE}_{rt} + \varepsilon_{it}, \end{aligned} \tag{3.3}$$

Table 3.7 shows that the coefficient on the interaction between log international gold prices and artisanal gold mining areas is positive, but it is statistically significant only in explaining the increase in battle-related fatalities. Furthermore, while the coefficient on the interaction between artisanal gold mining areas and the post-refinery period becomes smaller in magnitude across different battle-related outcomes, it remains statistically significant throughout. This suggests that the refinery effect is not fully explained by international price fluctuations and reflects an additional, persistent income shock.

Table 3.7. Effect of Refinery Opening and Gold Price on Battle

Type of Conflict	Battle					
Dependent Var.	Pr(event > 0)		IHS Events Counts		IHS Fatalities	
	(1)	(2)	(3)	(4)	(5)	(6)
ASGM×PostRefinery	0.259^{*}	0.265^{*}	0.872^{**}	0.890^{**}	1.266^{**}	1.282^{***}
	(0.107)	(0.106)	(0.282)	(0.281)	(0.387)	(0.386)
ASGM×Dist×PostRefinery	-0.017	-0.017	-0.070^{**}	-0.071^{**}	-0.103^{**}	-0.103^{**}
	(0.011)	(0.011)	(0.026)	(0.026)	(0.034)	(0.034)
$\ln(\text{Price}_t)\times$ASGM	0.063	0.067	0.132	0.126	0.605^{*}	0.582^{*}
	(0.087)	(0.088)	(0.207)	(0.210)	(0.293)	(0.295)
Mean of Dep. Var.	0.098	0.096	0.173	0.171	0.195	0.193
Cell FE	Yes	Yes	Yes	Yes	Yes	Yes
Region-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Cells with New Mines	Yes	No	Yes	No	Yes	No
Observations	8500	8280	8500	8280	8500	8280
Adjusted R^2	0.449	0.456	0.682	0.690	0.562	0.570

3.8 Conclusion

This paper provides new evidence on the cross-border effects of natural resource policy by examining the impact of Uganda's first gold refinery, established in 2016, on conflict dynamics in artisanal and small-scale gold mining (ASGM) areas in eastern Democratic Republic of Congo (DRC). Using novel, high-resolution data on ASGM sites combined with georeferenced conflict data, we find a statistically significant and economically meaningful increase in the likelihood of conflict in affected areas following the refinery's establishment. According to our estimates, we observe a 12.8% increase in the likelihood of violent battles, a 37.5% increase in conflict events, and a 45.3% rise in battle-related fatalities in affected regions post-2016.

In terms of mechanisms, our results are consistent with the idea that the refinery increased incentives for gold smuggling from eastern DRC into Uganda. As it is a setting with weak state-capacity and limited property rights, the establishment of the refinery increase significantly the benefit for appropriation.

Our findings carry important policy implications. This study highlights the unintended consequences of development initiatives aimed at capacity building or structural transformation—often supported by international donors. While these projects are typically intended to promote economic growth and strengthen local value-added exports, in this case, they inadvertently exacerbated conflict in neighboring regions. The results represent a clear case of a “beggar-thy-neighbor” dynamic: while Uganda benefits from increased exports and investment, the DRC bears the social and human cost of heightened violence.

These insights are especially relevant today, as global demand for critical minerals surges in the context of green energy transitions. Although our analysis focuses on gold, the findings may extend to other minerals central to the artisanal and small-scale mining (ASM) sector. As countries compete to capture mineral resources, they may be incentivized to appropriate mineral sites when extraction occurs outside their territory. A more coordinated, transnational approach to regulating resource trade and refining is therefore crucial to prevent fueling conflict and undermining regional stability.

Appendix

3.A Data Sources

Table 3.A.1. Variable Definitions

Regression Var. (Source)	Description
ASGM (IPIS)	A binary variable equal to 1 if any artisanal gold mine surveyed by IPIS on or before 2016 falls within the geographic boundary of the grid cell. Artisanal mining sites are often clustered, and multiple surveyed mines may fall within the same grid cell.
Battles (ACLED)	Events involving sustained armed combat between organized armed groups. This includes direct confrontations between government forces, non-state actors, or between rival non-state groups Raleigh and Dowd (2015).
Violence Against Civilians (ACLED)	Incidents where armed groups target unarmed civilians—e.g., killings, abductions, and assaults. This category captures deliberate acts of violence by organized actors against non-combatants (Raleigh and Dowd, 2015).
Temperature (ERA5)	(Desmet et al., 2024) computed the annual mean temperature for each cell using monthly data from the ERA Reanalysis dataset (Muñoz-Sabater et al., 2021). Standardized yearly deviations are used for comparability.
Precipitation (ERA5)	Monthly precipitation data from ERA5 is seasonally standardized (1989–2020), summed by year, and then standardized again to obtain cell-level annual rainfall deviations (Desmet et al., 2024).
Deforestation (Hansen et al., 2013)	Expressed as the percentage of each grid cell's area identified as forest loss. Forest cover is vegetation >5 meters tall. Pixels are classified as initially forested if they had 25% cover in the earliest period (Girard et al., 2022).

3.A.1 IPIS Surveys

The locations of artisanal mines in Eastern DRC come from the georeferenced survey data collected by the International Peace Information Service (IPIS).[16] IPIS has been conducting field-based surveys of artisanal mining sites in Eastern DRC since 2009. The version used in this study was downloaded in 2024 and includes data collected up to 2023. The unit of observation in the IPIS dataset is the mining site. Surveyor teams visit artisanal and small-scale mining (ASM) sites in the field, meet with miners, and complete predefined questionnaires using mobile phones. The data provide coordinates, the year of site visits, the types of minerals extracted, the presence of armed actors, and other characteristics of artisanal mining sites.

Regarding geographic coverage and representativeness, IPIS "strive[s] for [its] dataset to be as representative as possible of the small-scale mining sector in eastern DRC." However, the data is not intended to be representative of artisanal and small-scale mines across the entire DRC, given the highly localized and mineral-specific nature of mining and conflict dynamics.[17] Instead, IPIS states that its data is "relatively complete and representative" within the principal provinces it surveys, such as North and South Kivu. In these conflict-affected areas, artisanal gold mining is often intertwined with local armed group activity. This makes the IPIS dataset an appropriate and credible source available for studying the relationship between artisanal gold mining and violence in Eastern DRC.

Site selection in the IPIS dataset is guided by a combination of priorities: broad coverage across Eastern DRC (including remote areas), repeated visits to key sites with recent changes, and, in some cases, donor-specific requests (e.g., focus on certain minerals or validated sites). Survey coverage is also shaped by security conditions. In some years (e.g., 2013–2014), IPIS "were not able to visit all of the planned mine sites because of security."

We summarize the accessibility of artisanal gold mining sites using supplementary information documented by IPIS surveyors during each site visit. A total of 1,673 artisanal gold mines were surveyed on or before 2016. Of these, 96.8% were visited in person, while the remaining 3.2% were

[16] The data can be accessed at https://ipisresearch.be/home/maps-data/open-data/ and the technical notes on survey design can be found at https://ipisresearch.be/home/maps-data/open-data/open-data-faq-dictionary/.

[17] For example, small-scale diamond mines in the Kasaï provinces are not covered by IPIS.

surveyed remotely (e.g., by phone). Among the 36 remotely surveyed sites, 27 explicitly cited insecurity as the reason for not conducting an on-site visit. In addition, over 94% of the sites were geolocated using GPS, with only a small share recorded through approximate coordinates.

Because survey coverage varies across years due to changes in security conditions, there is a risk of endogeneity between conflict intensity and the timing of site visits. To mitigate this concern, we abstract from the temporal dimension of the IPIS data and focus only on the geographic distribution of mining sites and the type of mineral extracted. That is, we treat the spatial pattern of artisanal gold mining as time-invariant for the purpose of our analysis, using all available site coordinates that were surveyed on or before 2016, irrespective of the exact year of the visit. A grid cell is classified as an artisanal gold mining cell if it contains at least one surveyed gold mine. Due to the clustered nature of mining activity along gold belts, the number of artisanal gold mines per cell ranges from one to eight.

3.A.2 Distance to Refinery

To compute travel distance, we begin by importing road network data for the African continent from OpenStreetMap into QGIS and correcting geometry errors using the “Fix Geometries" tool. The road network is then clipped to our study area, defined by the merged administrative boundaries of the Democratic Republic of Congo (DRC) and Uganda. To ensure consistency and relevance, we retain only functional road segments by excluding features tagged as “waterway" or “aerialway" and filtering to include only relevant “highway" types (e.g., primary, secondary, tertiary, trunk, residential). A similar filtering process is applied to the Uganda road data. The cleaned road networks from both countries are then merged, and we apply a topological correction (using `v.clean` with a 0.05-degree tolerance) to ensure that adjacent road segments are properly connected. We cross-validate the cleaned road network against Google Earth, Google Maps, and the ESRI Transportation basemap, paying particular attention to the presence of rivers, and verifying whether bridges exist where road segments appear to cross waterways.

Using the `geopandas` and `networkx` libraries in Python, we load the cleaned and connected road network and reproject all spatial data into a consistent and suitable projected coordinate system. Centroids are computed for each grid cell and used as origin points, while the location of Entebbe International Airport

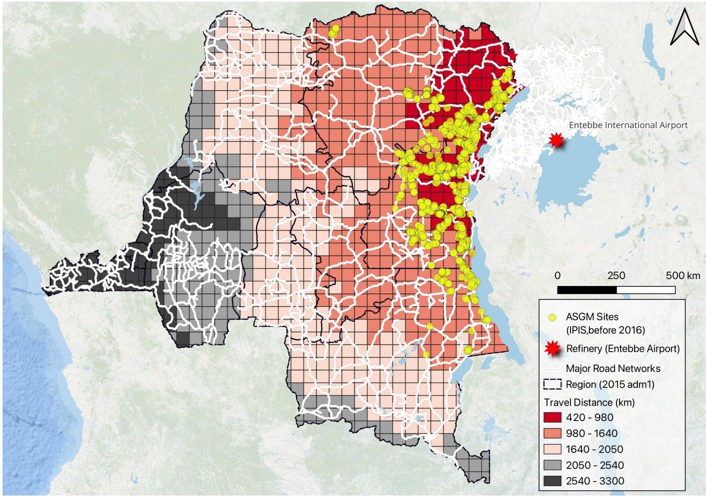


Figure 3.A.1. Travel Distance to Ugandan Gold Refinery

Notes: This figure provides an overview of travel distances to the gold refinery near Entebbe, the distribution of artisanal gold mines, the road network used to compute travel distances, and the provincial boundaries of the DRC as of 2015. The first two distance intervals: (i) 420–980 km and (ii) 980–1640 km roughly correspond to grid cells below and above the median distance to the refinery, which we refer to as “near" and “far" ASGM cells, respectively.

Source: OpenStreetMap, Google Map, ESRI and ESRI Basemap

serves as the destination. Each point is snapped to its nearest location on the road network to ensure valid path construction. We then construct a network graph and compute the shortest travel distances from each grid cell centroid to the airport along the road network. We focus exclusively on travel distance and abstract from variation in travel speed across road types.

3.B Additional Results and Robustness Checks

3.B.1 Violence Against Civilians

This section presents results using violence against civilians (VAC) as the outcome variable. Figure 3.B.1 and Figure 3.B.2 display event study estimates based on equation (3.2), while Table 3.B.1 reports the corresponding coefficients in tabular form. These results mirror the main analysis using battle-related conflict outcomes. The event study captures both the dynamic treatment effects and the heterogeneous impact of the gold refinery opening on VAC, depending on whether artisanal gold mining cells are located closer to or farther from the refinery in terms of travel distance.

In addition, Table 3.B.1 reports the effect of the gold refinery opening on violence against civilians, accounting for spatial and temporal correlation using Conley-Hsiang standard errors. This mirrors the robustness analysis presented in Table 3.6, which uses battle-related conflict as the outcome. The results remain robust, and the overall conclusions are unchanged.

Finally, Table 3.B.3 presents the results of a robustness check that includes international gold price shocks in the analysis of violence against civilians. The findings suggest that while rising international gold prices are associated with increases in both the frequency (event counts) and intensity (fatalities) of violence against civilians, the coefficients on the price interactions are not statistically significant. Importantly, our main results highlighting the post-refinery effect in artisanal gold mining areas and the heterogeneous impact by travel distance remain robust.

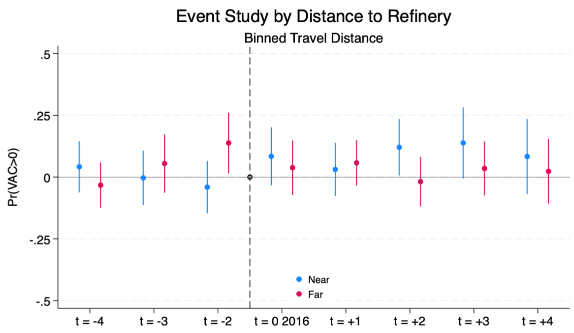


Figure 3.B.1. Event Study of Violence Against Civilians Likelihood by Distance to Refinery

Notes: This figure plots the coefficients and associated 95% confidence intervals estimated using equation (3.2). The dependent variable is binary and equals one if at least one violent event against civilians occurred in the grid in a given year, and zero otherwise. Distance to the gold refinery is divided into "near" and "far" based on whether an ASGM grid cell falls below or above the median distance.

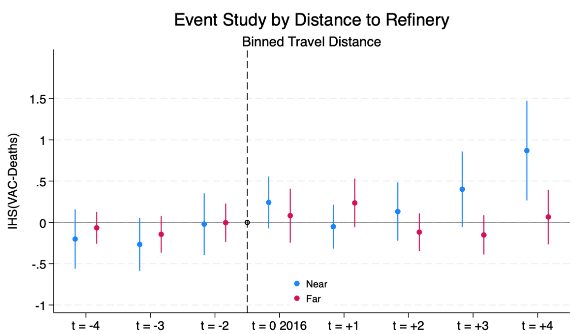


Figure 3.B.2. Event Study of Deaths from Violence Against Civilians by Distance to Refinery

Notes: This figure plots the coefficients and associated 95% confidence intervals estimated using equation (3.2). The dependent variable is the inverse hyperbolic sine (IHS) transformation of deaths resulting from violence against civilians. Distance to the gold refinery is divided into "near" and "far" based on whether an ASGM grid cell falls below or above the median distance.

Table 3.B.1. Tests for the Parallel Trends Assumption (Outcome: Violence Against Civilians)

	(1)	(2)	(3)	(4)
Event Time	Pr(VAC>0)Far	Pr(VAC>0)Near	IHS(VAC)Far	IHS(VAC)Near
4 years before	-.047 (.052)	-.015 (.054)	-.066 (.099)	-.201 (.183)
3 years before	.076 (.062)	.07 (.064)	-.144 (.114)	-.266 (.164)
2 years before	.012 (.066)	.017 (.068)	-.003 (.119)	-.021 (.19)
Year Refinery	-.02 (.057)	.138 (.066)*	.082 (.167)	.242 (.16)
1 year after	.166 (.07)**	.051 (.051)	.236 (.151)	-.052 (.136)
2 years after	.039 (.06)	.145 (.075)*	-.118 (.116)	.131 (.18)
3 years after	.004 (.047)	.156 (.071)*	-.152 (.121)	.402 (.233)*
4 years after	.154 (.067)*	.205 (.07)**	.066 (.169)	.869 (.308)**

Notes: $^{*}\ p < 0.05$, $^{*}* \ p < 0.01$, $^{*} * * \ p < 0.001$; Tabulated Results of Figure 3.B.1 and Figure 3.B.2. Robust standard errors clustered at the grid-cell level are reported in parentheses. The dependent variable is either a binary indicator equal to 1 if any conflict event (violence against civilians) occurred in a given year, or a continuous non-negative count (e.g., number of events or fatalities), to which we apply the inverse hyperbolic sine (IHS) transformation. This allows us to approximate a log transformation while retaining zero values and estimate all models within a unified linear framework. Columns (1) and (2), as well as columns (3) and (4), are estimated jointly using equation (3.2). “Near" and “Far" refer to grid cells located below and above the median distance to the border, respectively.

Table 3.B.3. Robustness Check: Including International Gold Prices (Outcome: Violence Against Civilians)

Type of Conflict	Violence Against Civilians					
Dependent Var.	Pr(event > 0)		IHS Events Counts		IHS Fatalities	
	(1)	(2)	(3)	(4)	(5)	(6)
ASGM×PostRefinery	0.250^{*}	0.251^{*}	0.839^{**}	0.837^{**}	1.043^{**}	1.056^{**}
	(0.119)	(0.119)	(0.299)	(0.299)	(0.359)	(0.359)
ASGM×Dist×PostRefinery	-0.020	-0.020	-0.078^{**}	-0.078^{**}	-0.084^{**}	-0.084^{**}
	(0.011)	(0.011)	(0.027)	(0.027)	(0.031)	(0.031)
$\ln(\text{Price}_t)\times$ASGM	-0.090	-0.088	0.102	0.123	0.337	0.327
	(0.115)	(0.116)	(0.265)	(0.267)	(0.321)	(0.325)
Mean of Dep. Var.	0.106	0.104	0.181	0.179	0.145	0.145
Cell FE	Yes	Yes	Yes	Yes	Yes	Yes
Region-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Cells with New Mines	Yes	No	Yes	No	Yes	No
Observations	8500	8280	8500	8280	8500	8280
Adjusted R^2	0.478	0.484	0.670	0.677	0.523	0.527

Notes: $^{*}\ p < 0.05$, $^{*}{*}\ p < 0.01$, $^{*}{*}{*}\ p < 0.001$; Robust standard errors clustered at the grid-cell level are reported in parentheses. Distance is expressed in units of 100 kilometers. Columns (1), (3), and (5) use equation (3.3). Columns (2), (4), and (6) exclude grid cells where ASGM sites were first recorded after 2016. This tests whether gold price fluctuations explain conflict variation.

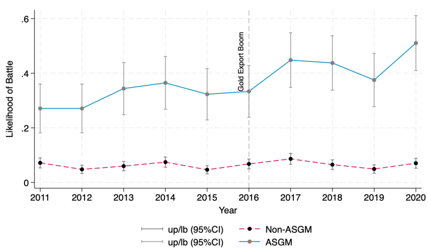
3.B.2 Descriptive Trends

This section presents descriptive trends in conflict-related outcomes before and after the opening of the gold refinery in 2016, comparing artisanal and small-scale gold mining (ASGM) areas with non-ASGM areas. These figures provide a visual assessment of whether pre-treatment trends were parallel, which is important for evaluating the validity of the identification strategy. Figure 3.B.3 presents the descriptive trends for battles, while Figure 3.B.4 shows the corresponding trends for violence against civilians.

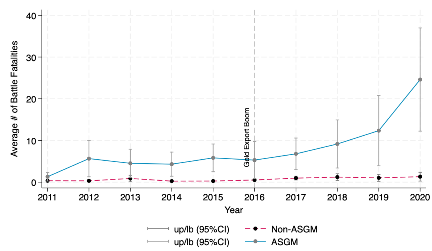
Table 3.B.4. Effect of Gold Refinery Opening on Violence Against Civilians and Descriptive Trends

Type of Conflict	Violence Against Civilians					
Dependent Var.	Pr(event > 0)		IHS Events Counts		IHS Fatalities	
	(1)	(2)	(3)	(4)	(5)	(6)
ASGM×PostRefinery	0.244*	0.245*	0.846*	0.845*	1.066*	1.079*
	(0.117)	(0.118)	(0.359)	(0.356)	(0.466)	(0.462)
ASGM×Dist×PostRefinery	-0.020*	-0.020*	-0.078*	-0.078*	-0.084*	-0.084*
	(0.010)	(0.010)	(0.033)	(0.033)	(0.041)	(0.040)
Cell FE	Yes	Yes	Yes	Yes	Yes	Yes
Region-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Cells with New Mines	Yes	No	Yes	No	Yes	No
Observations	8500	8280	8500	8280	8500	8280

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Robust standard errors clustered at the grid-cell level are reported in parentheses. Distance is expressed in units of 100 kilometers. The IHS transformation allows log-like modeling for variables with zero values.



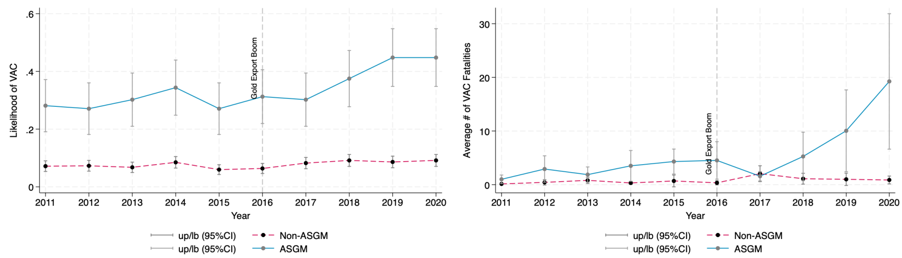
(a) Likelihood (Battle)



(b) Fatalities (Battle)

Figure 3.B.3. Descriptive Trends in Battle Likelihood and Fatalities: ASGM and Non-ASGM Areas

Notes: ACLED data are aggregated by computing the mean and standard error of a binary indicator and fatalities. Means are plotted with 95% confidence intervals.



(a) Likelihood (Violence Against Civilians) **(b)** Fatalities (Violence Against Civilians)

Figure 3.B.4. Descriptive Trends in VAC Likelihood and Fatalities: ASGM and Non-ASGM Areas

Notes: The ACLED data are aggregated by computing the mean and standard error of two variables: (i) a binary indicator equal to 1 if a battle event occurred and 0 otherwise, and (ii) the number of reported fatalities. These statistics are calculated separately for ASGM and non-ASGM grid cells, and the resulting means are plotted with 95% confidence intervals.

3.B.3 Newly Surveyed Artisanal Gold ines

To address the concern that a positive income shock in artisanal gold mining, which affects a labor-intensive sector, could theoretically activate the opportunity cost channel and reduce violence, we conduct additional analysis to better understand the plausibility of this mechanism in our context.

While the affected population is substantial, the opportunity cost channel may only be relevant in areas where new mining opportunities arise and labor can be redirected away from conflict. In contrast, in pre-existing mining areas already under contested control of armed actors, the increased value of gold may reinforce incentives for appropriation. To examine this distinction, we use an alternative treatment definition that includes all artisanal gold mining sites, regardless of whether they were identified before or after the refinery opened.

Under this broader definition, the estimated coefficients remain positive but are smaller in magnitude compared to our baseline results. This suggests that the conflict-intensifying impact of the refinery is more pronounced in areas with pre-existing mining activity, consistent with the idea that appropriation incentives are stronger where artisanal mining was already established. In contrast, in newly established mining areas, the opportunity cost channel may dominate—reducing violence as individuals shift toward income-generating activities.

One major newly established mining area is located in the far northeastern region of the DRC, near the border with Uganda and the South Sudan (See the map shown in Figure 3.B.5). This area had previously experienced attacks by the Lord's Resistance Army (LRA), but the group's influence has declined significantly since the mid-2010s. This stands in sharp contrast to the neighboring Ituri province, which is also close to Uganda but has experienced a very different conflict trajectory, with ongoing violence and contestation over resources.

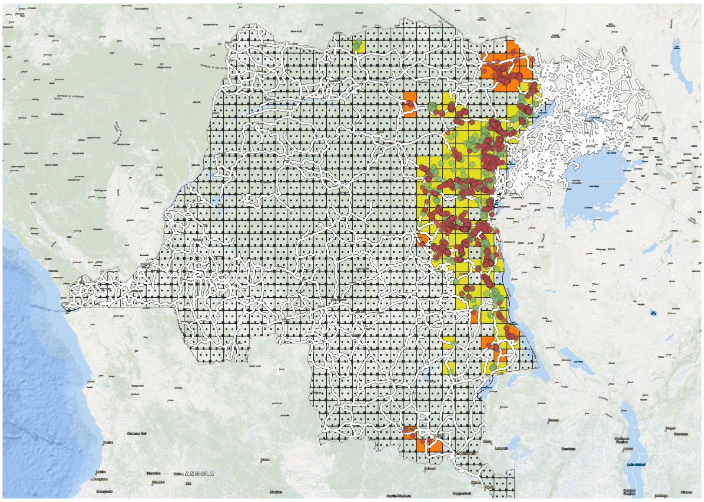


Figure 3.B.5. Locations of Newly Surveyed Artisanal Gold Mines

Notes: This map displays the locations of newly surveyed and pre-existing artisanal gold mining sites. Green dots indicate sites surveyed in 2016 or earlier, while red dots represent sites surveyed after 2016. Yellow grid cells correspond to areas with pre-existing artisanal gold mining activity, and orange grid cells indicate newly surveyed ASGM areas. A grid cell is classified as a *newly surveyed ASGM area* only if all mining sites within the cell were added to the survey *after* 2016. That is, no site in the cell was recorded in or before 2016.
Source: IPIS (2009 to 2023), OpenStreetMap, and ESRI Basemaps

Table 3.B.5. Robustness Check: Alternative Treatment Definition (Outcome: Battle)

Type of Conflict	Battle					
Dep. Var.	Pr(event > 0)		IHS Events Counts		IHS Fatalities	
	(1)	(2)	(3)	(4)	(5)	(6)
ASGM×PostRefinery	0.264*		0.881**		1.308***	
	(0.107)		(0.282)		(0.388)	
ASGM×Dist×PostRefinery	-0.017		-0.071**		-0.103**	
	(0.011)		(0.026)		(0.034)	
ASGMAll×PostRefinery		0.152		0.581*		0.824**
		(0.081)		(0.214)		(0.303)
ASGMAll×Dist×PostRefinery		-0.007		-0.042*		-0.056*
		(0.007)		(0.018)		(0.025)
Mean of Dep. Var.	0.098	0.098	0.173	0.173	0.195	0.195
Cell FE	Yes	Yes	Yes	Yes	Yes	Yes
Region-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8500	8500	8500	8500	8500	8500
Adjusted R^2	0.449	0.449	0.682	0.681	0.562	0.560

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Robust standard errors clustered at the grid-cell level are in parentheses. Distance is in 100 km. Columns (1), (3), and (5) use the baseline ASGM treatment; columns (2), (4), and (6) use an extended treatment including all ASGM sites ever recorded between 2009 and 2023.

Table 3.B.6. Robustness Check: Alternative Treatment Definition (Outcome: Violence Against Civilians)

Type of Conflict	Violence Against Civilians					
Dependent Var.	Pr(event > 0)		IHS Events Counts		IHS Fatalities	
	(1)	(2)	(3)	(4)	(5)	(6)
ASGM×PostRefinery	0.244*		0.846*		1.066*	
	(0.120)		(0.300)		(0.362)	
ASGM×Dist×PostRefinery	-0.020*		-0.078*		-0.084*	
	(0.011)		(0.027)		(0.031)	
ASGMAll×PostRefinery		0.138		0.512*		0.728**
		(0.096)		(0.232)		(0.280)
ASGMAll×Dist×PostRefinery		-0.011		-0.047*		-0.055*
		(0.008)		(0.020)		(0.023)
Mean of Dep. Var.	0.106	0.106	0.181	0.181	0.145	0.145
Cell FE	Yes	Yes	Yes	Yes	Yes	Yes
Province-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8500	8500	8500	8500	8500	8500
Adjusted R^2	0.478	0.478	0.670	0.668	0.522	0.521

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Robust standard errors clustered at the grid-cell level are reported in parentheses. Distance is expressed in units of 100 kilometers. Columns (1), (3), and (5) are estimated using equation (3.1). Columns (2), (4), and (6) use an alternative treatment definition, based on all artisanal gold mines ever recorded between 2009 and 2023 from survey data.

4

Economics or Culture? Social Identification in Sub-Saharan Africa

This paper examines how the economic conditions of an individual's ethnic group and its cultural distance from the nation influence social identification in sub-Saharan Africa. I construct a novel measure of economic conditions at the ethnic group level using a shift-share design that captures the agricultural value of each group's historical homeland. To measure cultural distance, I rely on the linguistic distance of each ethnic group relative to the national average. Focusing on ethnic versus national identification, I find two main results. First, individuals who are more linguistically distinct from the rest of the country are more likely to identify with their ethnic group. Second, as the economic conditions of an ethnic group improve, individuals tend to identify more strongly with their group. Taken together, these results provide empirical support for the cost-benefit framework of identity formation in the political economy literature and underscore the role of both economic incentives and cultural distance in shaping social identification.

4.1 Introduction

Since the seminal work of Akerlof and Kranton (2000), it has been widely recognized that how individuals identify themselves is crucial to understanding economic outcomes. More recently, scholars have emphasized that people's identity also plays a central role in the political environment (Ansolabehere and Puy, 2016; Grossman and Helpman, 2021). Much of this literature builds on the conceptual framework proposed by Shayo (2009), which posits that people identify with social groups based on a trade-off between the cognitive costs of group membership and the benefits of belonging to high-status groups. While individuals tend to align themselves with high-status groups, the cognitive cost of identification increases when those groups are perceived as culturally distant. Although this trade-off has been extensively explored in theory, there is limited empirical evidence on how it operates in practice. This study contributes to filling this gap by demonstrating how the economic status of social groups and cultural distance from them influence individual social identification.

Two main empirical challenges arise in this analysis: obtaining an exogenous measure of group economic status and quantifying how much cultural distance they are from a reference group. To address these challenges, I focus on sub-Saharan Africa and investigate how these two forces influence ethnic versus national identification. To measure ethnic group economic status, I compute the agricultural value of the ethnic group's homeland, leveraging the fact that agriculture constitutes a large share of household production and consumption in Africa compared to other regions. For cultural distance, I use language as a salient dimension of identity in sub-Saharan Africa and calculate the linguistic distance of each ethnic group relative to its country. My findings support the intuition from political economy literature that both economic status and cultural distance shape how individuals categorize themselves into groups.

My approach proceeds as follows. First, I use the Afrobarometer survey to measure individuals' social identification with their ethnic group versus the nation. Second, since the survey includes information on the language spoken in the household, I link each household to its corresponding ethnolinguistic group. Third, I employ a shift-share design to proxy the economic status of each ethnic group by estimating the agricultural value of its historical homeland. This combines high-resolution, time-invariant spatial data on cash crop production with international cash crop prices.[1] This provides a novel way to capture economic conditions at the ethnic group level. While similar shift-share approaches have been used in the literature to measure economic conditions at the grid-cell level, this study uniquely focuses on shocks at the ethnic group level. [2] Finally, to measure cultural distance, I compute the ethnic distance of each ethnic group from the rest of the country using a population-weighted average linguistic distance, following a method similar to Desmet et al. (2024).

I find two main results. First, individuals belonging to ethnolinguistically distant groups, relative to the rest of their country, are more likely to identify with their ethnic group. I interpret linguistic distance as a cultural cost: when the perceived distance from the nation is large, the cost of national identification increases, making individuals more likely to align with their ethnic group. Second, as the economic conditions of ethnic groups improve, individuals are more likely to identify with their group, suggesting that higher group status increases the perceived benefits of group affiliation. When examining the interaction between economic status and linguistic distance, I find that the effect of economic status is strongest among individuals from more linguistically distant groups. Taken together, these results provide empirical support for the cost-benefit framework of social identification proposed by Shayo (2009).

The strength of my empirical strategy lies in the exogeneity of changes in ethnic groups' economic status, as fluctuations in international prices are unlikely to be correlated with unobserved factors that influence social identification. Additionally, the strategy incorporates a rich set of fixed effects to address potential threats to causal inference. Ethnic group-country fixed effects control for time-invariant unobserved heterogeneity across ethnic groups, since the identifying variation comes from within-group changes over time. Country-year fixed effects account for national shocks or policies that affect all ethnic groups within a country in a given year. I also include grid cell fixed effects, which control for local differences such as access to public

[1] The focus on cash crops is important because price fluctuations can differentially impact the status of an ethnic group depending on whether the crop is used primarily for subsistence or for sale in international markets.

[2] To the best of my knowledge, the only other study that examines shocks at this level is Berman et al. (2023), who analyze a mining shock affecting ethnic groups but find no significant effect on group-level welfare.

services or proximity to the capital—factors that could influence identification. Finally, I control for individual-level characteristics, including profession, age, gender, and whether the respondent lives in an urban or rural area.

My work builds on a growing literature on the endogenous formation of social identities. Empirically, several studies have linked the strength of ethnic identification to factors such as whether the ethnic group holds political power, the level of electoral competition, or perceptions of relative deprivation (Green, 2020; Eifert et al., 2010; Berman et al., 2023). In contrast, other research has associated stronger national identification with shared collective experiences or residence in more prosperous regions (Depetris-Chauvin et al., 2020; Ahlerup et al., 2017). However, these studies largely overlook how individuals respond when the economic conditions of their social group change. My work addresses this gap by providing robust causal evidence that improvements in an ethnic group's economic status can lead to changes in individual social identification.

My results also contribute to an emerging strand of literature on the consequences of cultural distance. Guarnieri and Tur-Prats (2023) demonstrate that cultural distance in gender norms between perpetrators and victims influences the likelihood of sexual violence. Guarnieri (2024) show that greater cultural distance between ethnic groups and the central government increases the likelihood of conflict over political power. Similarly, Desmet et al. (2024) find that a location's ethnic remoteness—defined as its population-weighted average ethnic distance to the rest of the country—acts as a barrier to participation in local trade networks. I contribute to this literature by showing that cultural distance also plays a significant role in shaping individual social identification.

Furthermore, this paper also provides new insights into the literature on the determinants of secessionist sentiment by providing new insights about conditions under which support for secession emerges. Using a natural experiment, Gehring and Schneider (2020) show that oil discoveries along the Scottish coast increased regional income and, in turn, support for secessionist parties. Desmet et al. (2025) examine whether secessionist demands in subnational regions worldwide are driven more by differences in income per capita or by ethnolinguistic identity. They find that while both factors matter, identity emerges as a sufficient condition for secessionist sentiment, often overriding economic considerations. Similarly, Álvarez Pereira et al. (2018) find that wealthier regions are more supportive of secessionist parties, but only when they also possess a distinct cultural identity. My paper builds on this literature but departs from it by shifting the level of analysis from regions to individuals. Whereas previous work typically uses aggregate regional data, I use individual-level survey data, combined with group-level measures of income and linguistic distance. This allows for a more granular understanding of how individuals can identify with their groups.

My paper contributes to the understanding of identity formation by illustrating how ethnic group economic status and cultural distance influence individual identification with ethnic versus national identities in sub-Saharan Africa. This study sheds light on two key components emphasized in the recent literature on the political economy of identity about how identity affects political and economic outcomes, and in particular how the rise of political parties that exploit identity-based divisions can be understood through these mechanisms (see Guriev and Papaioannou (2022) for a review).

This paper is organized as follows: Section 2 presents the conceptual framework underlying the analysis. Section 3 provides an overview of the data and the various measures employed. Section 4 outlines the empirical strategy. Sections 5 and 6 present the main results and some robustness checks. Section 7 discusses the results and Section 8 concludes.

4.2 Conceptual Framework

Social identity is defined by social psychologists as "the individual's knowledge that he belongs to certain social groups together with some emotional and value significance to him of the group membership" (Tajfel, 1974). Based on this concept, Tajfel (1981) and Tajfel (1979) developed Social Identity Theory (SIT), which posits that a significant part of an individual's self-esteem is derived from a sense of belonging to social groups. SIT emphasizes that individuals enhance their self-image through favorable comparisons between their in-group and relevant out-groups. A further development of SIT, known as Self-Categorization Theory (SCT), was introduced by Turner (1987). SCT shifts the focus from intergroup comparisons to the process by which individuals categorize themselves within a group. It highlights that identification depends not just on membership but also on the degree to which individuals perceive themselves as aligning with the group prototype. According to SCT, this identification process involves both a cognitive mechanism—where individuals classify themselves based on perceived group similarities and distinctions—and an affective component, reflecting the emotional value they attach to group membership.

Building on the foundational ideas of Social Identity Theory (SIT), Akerlof and Kranton (2000, 2010) introduced the concept of identity into economics by defining it as “a person’s sense of self” and incorporating a self-esteem component into the utility function. Later, Shayo (2009) developed a formal framework that more explicitly integrates insights from Self-Categorization Theory (SCT). In this model, individuals identify with groups based on a trade-off between the benefits of belonging to a high-status group and the cognitive costs associated with dissimilarity from the group’s prototypical member. Shayo applied this framework to explain why individuals may support political parties that do not align with their material or economic interests. Since then, the framework has been extended in the political economy literature to examine how identity shapes voter preferences, how shifts in social identification influence policy views, and how identity affects information processing (Ansolabehere and Puy, 2016; Grossman and Helpman, 2021; Karakas and Mitra, 2019).

In this framework, I use economic status as a proxy for the benefits of group identification, measured by the agricultural value of each ethnic group’s historical homeland. I exploit within-group variation over time to examine whether, in years when a group experiences higher economic gains—that is, when its economic status increases—feelings of ethnic belonging also intensify. To capture the costs associated with identity formation, I use the nation as the reference group and measure cultural distance as the linguistic distance between each ethnic group and the population-weighted average ethnic group in the country. The underlying hyphotesis is that groups that are linguistically distant from the national average face higher cognitive costs when identifying with the nation.

4.3 Data and Measurement

This section describes the data and outlines the construction of the main variables used in the paper. It also discusses their validity in relation to the cost-benefit framework of social identification.

4.3.1 Individual identification

To measure individual social identity, I use data from the Afrobarometer survey, which provides repeated cross-sectional data across multiple waves. I focus on waves 2 through 8, covering a total sample of more than 100,000 households collected between 2002 and 2021. The survey contains information on attitudes toward institutions, perceptions of poverty and violence, and a range of individual characteristics, including occupation, age, gender, education, and the household's primary language. Panel (a) of Figure 4.1 shows the geographic distribution of surveyed households.

My main outcome variable captures an individual's identification with the nation versus their ethnic group. It is derived from responses to the following question: *Let us suppose that you have to choose between being a [national identity] and being a [respondent's identity group]. Which of these two groups do you feel most strongly attached to?.* Household can answer the following: *1=I Feel Only (r's group), 2=I Feel More (r's group) than [Nation], 3=I Feel Equally [Nation] and (r's groups), 4=I Feel More [Nation] than (r's groups), 5=I feel only [Nation].* I construct a binary variable equal to 1 if the respondent identifies solely with their ethnic group or more strongly with their ethnic group than with their nation, and 0 otherwise. This coding follows a common approach in the literature using Afrobarometer data to measure social identity (Rohner et al. (2013); Depetris-Chauvin et al. (2020); Berman et al. (2023)).

As discussed in Section 4.4, my analysis exploits within-ethnic group variation over time. To ensure sufficient longitudinal variation, I restrict the sample to ethnic groups that are consistently observed across all survey rounds.

4.3.2 Ethnic group status

The first step in calculating the economic status of each ethnic group is to link Afrobarometer households to their corresponding ethnolinguistic groups. To accomplish this, I use the mapping developed by Desmet et al. (2024), which matches Afrobarometer households to ethnolinguistic groups in the Ethnologue based on the primary language spoken in the household. This mapping is available for the following countries: Botswana, Ghana, Lesotho, Malawi, Mali, Namibia, Nigeria, South Africa, Tanzania, Uganda, Zambia, and Zimbabwe.

To identify the homelands of these ethnic groups, I use the World Language Mapping System (WLMS), which maps nearly 7,000 linguistic groups worldwide. These maps correspond to the 17th edition of the Ethnologue database Lewis et al. (2014). Using each ethnolinguistic group's homeland from the WLMS, I compute the economic status of each ethnic group with a shift-share design, where the *share* is derived from high-resolution spatial data on crops, and the *shift* is based on changes in international price data. The spatial crop data is sourced from the M3-Cropland dataset developed by Ramankutty et al. (2008), and the prices come from from the World Bank Commodity Price Data and the International Finance Statistics (IMF) Commodity Prices series.[3]

For each ethnic group e in country c at year t, economic status is defined as:

$$\text{Status}_{ect} = \left[\sum_{x=1}^{N} \Big(P_{xt} \times S_{ecx} \Big) \right] \tag{4.1}$$

where S_{ecx} represents the share of cash crop x in the homeland of ethnic group e within country c for crops with available international prices, and P_{xt} denotes the price of crop x in year t. Prices are normalized to 100 for the year 2000. Since changes in economic status over time are driven solely by fluctuations in international prices, this measure can reasonably be assumed to be exogenous to local conditions that might also influence social identification.

The focus on cash crops is guided by the observation that increases in international prices of staple crops are unlikely to significantly improve the economic conditions of ethnic groups, as these crops are primarily consumed locally rather than sold for income. I define cash crops as those contributing less than 1% to a country's total calorie consumption.[4] Food consumption data are obtained from the FAO Food Balance Sheets, and time-invariant consumption shares are calculated using average values from the 1985–2013 period. This approach follows the methodology of McGuirk and Burke (2020).

Does my status change the economic conditions of the group? A key element of this study is that my measure of economic status influences income at the ethnic group level. To test this, I build on the pioneering work of Henderson et al. (2012), and I use nighttime lights data measured by satellites as a proxy for income. I aggregate the nighttime lights data at the ethnic group

[3] World Bank Data: https://www.worldbank.org/en/research/commodity-markets; IMF data: https://www.imf.org/en/Research/commodity-prices.

[4] Table A1 show the classification of cash crops by each country.

homeland level using spatial GIS techniques.

To study the validity of my measure of status on income at ethnic group level, I use the following specification:

$$\text{Luminosity}_{ect} = \alpha_0 + \alpha_1 \text{Status}_{ect} + \beta_{ct} + \beta_{ec} + \epsilon_{ect} \tag{4.2}$$

where e denotes ethnic group, c denotes country, and t denotes year. Luminosity_{ect} captures income at the ethic group homeland, proxied by nighttime luminosity, for ethnic group e in country c at time t. Status_{ect} represents the economic status of ethnic e in country c at time t as defined in equation 4.1. β_{ct} denotes country-year fixed effects, and β_{ec} represents ethnic group-country fixed effects. I adjust the standard errors for spatial correlation within 500 km and for serial correlation following Conley (1999) and Hsiang (2010).

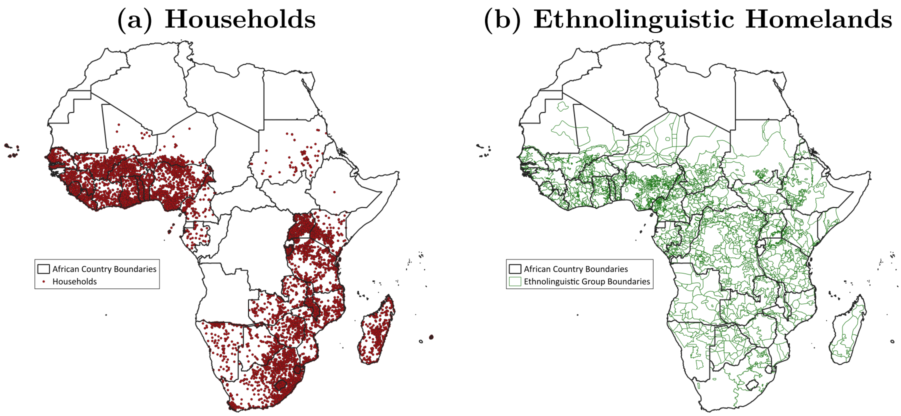
Table 4.1 presents the results from estimating Equation 4.2, showing that my measure of economic status is positively associated with income at the ethnic group level. This provides evidence that the measure is a valid proxy for capturing the economic conditions of ethnic groups. The results are based on countries where I am able to match Afrobarometer households to Ethnologue ethnolinguistic groups. However, the findings remain consistent when the analysis is extended to include all ethnic groups across sub-Saharan Africa.

Table 4.1. Economic Status and Income at Ethnic Group Level

	(1) Nighttime Luminosity
Status	0.027***
	(0.009)
Observations	19005

Notes: Standard errors (in parentheses) corrected for spatial correlation within 500 km radius and for infinite serial correlation (Conley, 1999; Hsiang, 2010). The dependent variable is nighttime luminosity at the ethnic group homeland. Ethnicity-country and country-year fixed effects are included. Status is defined as in Equation 4.1 using cash crops. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 4.1. Ethnolinguistic Homelands and Households



Notes: Panel (a) plots the distribution of households from Afrobarometer, rounds 2 to 8. Panel (b) plots the map of ethnolinguistic homelands from Ethnologue. The distribution of ethnic groups is based on data from the World Language Mapping System.

4.3.3 Culture distance

In sub-Saharan Africa, language serves as a prominent marker of cultural identity. Consistent with a substantial body of research, I use linguistic distance as a proxy for cultural distance. The literature offers two primary approaches for calculating linguistic distance between ethnic groups and their respective countries. On one hand, Desmet et al. (2024) measure ethnic distance as the population-weighted average distance from each ethnic group to the country's average ethnic group, resulting in a time-invariant measure. On the other hand, Guarnieri (2024) calculate linguistic distance relative to the government, which varies with changes in leadership.

To avoid potential confounding factors, I adopt the approach of Desmet et al. (2024), which calculates ethnic distance relative to the country's average ethnic group rather than to the government. Prior research has documented the presence of ethnic favoritism in Africa, showing that individuals who share the president's ethnicity benefit from preferential outcomes, including improved educational attainment, increased road construction in their districts, and higher levels of nighttime luminosity (Franck and Rainer, 2012; Burgess et al., 2015; De Luca et al., 2018). Moreover, individuals who share the president's ethnicity are also more likely to report stronger ethnic identification (Green, 2020). An additional reason for adopting this measure is that calculating ethnic distance relative to the country's average ethnic group—rather than to the government—better reflects the cultural rather than political dimension of national identity, thereby aligning more closely with the concept this paper seeks to capture.

My measure is designed to approximate the cultural distance of each ethnic group relative to the rest of its country. I define the cultural distance of ethnic group e in country c as follows:

$$LD_{ec} = \sum_{m}^{nc} s_{ec} s_m d_{ecm}. \tag{4.3}$$

where s_{ec} is the share of the population in country c that belong to ethnic group e; s_m is the share of the population in country c that belongs to ethnic group m; nc represents the total number of ethnic groups in country c; and d_{ec} the linguistic distance between ethnic group e and m. Linguistic distance is computed based on the number of shared branches in the Ethnologue linguistic tree, as defined by:

$$d_{ecm} = 1 - \left(\frac{b_{ecm}}{b_{max}}\right)^{\delta} \tag{4.4}$$

where b_{ecm} is the number of shared branches between languages e and m, b_{max} is the maximum number of branches that two languages can share, and δ is a parameter capturing how rapidly distance declines as the number of common branches increases. Higher values of δ cause the distance to decline more sharply, and maintain relatively high distance values unless many branches are shared. In other words, when δ is higher, only strong linguistic connections (those which shared a greater number of branches) are treated as closed, and languages that are genealogically unrelated remain highly distinct (for further discussion, see Desmet et al. (2009b)).

In this project, I apply three different values of δ (0.5, 0.05, and 0.0025), which I refer to throughout the paper as "DistanceFearon", "DistanceDesmet", and "DistanceGomes" corresponding to the main studies that use each specification (Fearon, 2003b; Desmet et al., 2009b; Gomes, 2020), respectively. This variation allows me to examine how different weightings of linguistic distance—emphasizing deeper versus more superficial similarities—affect social identification.

Which linguistic cleavages matter more? To examine the main effect of cultural distance on social identification and determine which degrees of linguistic distance are most predictive, I estimate the following specification:

$$Y_{iectj} = \gamma_0 + \gamma_1 \text{Distance}_{ec} + \beta_{ct} + \beta_j + \epsilon_{iect} \tag{4.5}$$

where i denotes household, e denotes ethnic group, c denotes country, j cell, and t denotes year. [5] Y_{iect} represents the social identification of household i, as described in 4.3.1, $Distance_{ec}$ captures the linguistic distance of ethnic group e i country c with respect to its country, and β_j and β_{ct} are cell and country-year fixed effects, respectively. Standard errors are clustered at ethnic group country-year to maintain cosistency with the main specification in the empirical strategy. Results clustered at ethnic group level are showned in Table B2 in the Appendix Section 4.A.2

The results are presented in Table 4.2. They suggest that the greater the linguistic distance between an ethnic group and the average ethnic group in the country, the more likely individuals are to identify with their ethnic group rather than with the nation. This finding is consistent with the conceptual framework: the more linguistically distant individuals are from their country, the greater the cognitive cost of identifying with the nation. Additionally, Table 4.2 shows that linguistic distances computed using higher values of δ have a stronger effect.

Table 4.2. Cultural Distance and Individual Identification

	(1)	(2)	(3)	(4)	(5)	(6)
DistanceGomes	0.055***			0.053***		
	(0.016)			(0.020)		
DistanceDesmet		0.061***			0.057***	
		(0.017)			(0.021)	
DistanceFearon			0.101***			0.084***
			(0.021)			(0.025)
Cell FEs	Yes	Yes	Yes	Yes	Yes	Yes
Indv Controls	No	No	No	Yes	Yes	Yes
Observations	101301	101301	101301	69082	69082	69082

Notes: Standard errors (in parentheses) are clustered by ethnic group-country-year. The dependent variable is a binary indicator equal to 1 if the individual identifies more with their ethnic group than with their nation. Distance is defined as in Equation 2 at varying levels of aggregation: DistanceFearon corresponds to $\delta = 0.5$, DistanceDesmet to $\delta = 0.05$, and DistanceGomes to $\delta = 0.0025$. Columns (4)-(6) include specifications individual controls. Country-year and cell fixed effects are included. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[5] I divide sub-Saharan Africa into 55 × 55 km grid cells and assign each household to a cell based on its geolocation.

Importantly, I do not interpret these results as demonstrating a causal effect, since my linguistic distance is, by definition, time-invariant. However, because this measure is highly persistent over time and determined prior to the dependent variable, I interpret the estimates as indicating a strong correlation. Throughout the paper, I show the results using distances computed with the three different δ.

4.3.4 Descriptives

Section 4..2 in the Appendix presents summary statistics for the main variables described above. Table A2 reports descriptive statistics for the core variables used in the analysis. Most individuals identify more strongly with their nation than with their ethnic group: out of 101,358 individuals in the sample, only 18,041 report identifying more with their ethnic group than with the nation (see Table A4). Table A2 also confirms the intuition discussed in Section 4.3.3: linguistic distances calculated using higher values of δ capture broader cultural differences, as reflected in their higher mean values.

Table A2 also shows that the variable $Status_{ect}$ exhibits substantial variation across ethnic groups, with a high standard deviation and variance. To explore this further, Table A5 presents the distribution of this variable by percentiles, illustrating that it is highly right-skewed. To complete the picture, Table A6 reports the between and within variation in $Status_{ect}$, showing that most of the variation comes from differences between ethnic groups rather than within groups over time. These patterns underscore the importance of normalizing the variable in the main empirical specifications.

To explore descriptively whether high values of the variable $Status_{ect}$ are systematically associated with more identification with the ethnic group, I divide variable $Status_{ect}$ into quartiles and examine the mean and standard deviation of the binary identity dependent variable within each group. Table A3 provides descriptive evidence that individuals from higher-status ethnic groups tend to identify more strongly with their ethnicity. The standard deviation remains stable between the four groups.

Finally, Tables A7 and A8 present pairwise correlations among the variables. The results show that $Status_{ect}$ and $Distance_{ec}$ are not correlated, indicating that higher-status groups are not systematically more or less linguistically distant from the rest of the country. They capture distinct dimensions of ethnic group heterogeneity.

4.4 Empirical Strategy

To estimate the effect of economic status and cultural distance on individual social identification, I use the following baseline specification:

$$Y_{iectj} = \beta_0 + \beta_1 \text{Status}_{ect} + \beta_2 \text{Status}_{ect} \times \text{Distance}_{ec} + \mathbf{X'}_i \beta + \beta_{ct} + \beta_e + \beta_j + \epsilon_{iectj} \tag{4.6}$$

where i denotes the individual, e the ethnic group, c the country, j the cell associated with the household, and t the year. Y_{iectj} is a binary variable equal to 1 if individual i, who belongs to ethnic group e in country c and in cell j at time t, identifies more strongly with their ethnic group than with the nation. Status_{ect} represents the economic status of ethnic group e in country c at time t, as defined in Equation 4.1, and Distance_{ec} captures the linguistic distance between ethnic group e relative to its country c, as defined in Equation 4.3. $\mathbf{X'}_i$ includes a set of individual characteristics; β_{ct} are country-time fixed effects; β_j are cell fixed effects; β_e are ethnicity effects, and ϵ_{iectj} is the idiosyncratic error term. I cluster standard errors at the level of my treatment, ethnic group country-year.[6] The main coefficients of interest are β_1, which measures the effect of economic status on social identification, and β_2 which captures how the relationship between economic status and social identification changes with linguistic distances.

My empirical strategy relies on the assumption that changes in the economic status of ethnic groups are exogenous to changes in social identification. Specifically, I assume that variation in my measure of ethnic group economic conditions over time is driven by external factors (changes in international prices), rather than by endogenous responses to shifts in individuals' social identification. In addition, my measure of linguistic distance is based on the Ethnologue linguistic tree and reflects the historical composition of ethnic groups in each country, which was determined long before the time period covered by my dependent variable. This ensures that the estimated effects of economic status on ethnic identification are not confounded by reverse causality.

The inclusion of fixed effects helps control for unobserved factors that could bias the results. Country-year fixed effects account for factors specific to each

[6]Section 4.A.2 in the Appendix reports results with standard errors clustered at ethnic-group country level.

country in a given year, including the "status of the nation". This is particularly important, as national-level shocks can influence individuals' identification with the nation (Depetris-Chauvin et al., 2020). Ethnicity fixed effects control for time-invariant unobserved heterogeneity across groups, such as cultural norms, historical legacies, or shared traditions. In addition, local factors—such as the provision of public goods, urbanization, or distance to the capital—can shape how individuals perceive their government and, in turn, affect their national or ethnic identification. To account for these local dynamics, I include cell fixed effects, which capture unobserved spatial characteristics that may influence identification patterns. Finally, I control for individual-level characteristics, including profession, gender, income, and whether the respondent lives in a rural or urban area.

4.5 Main Results

Table 4.3 presents the baseline results of this study, reporting the estimation of equation 4.6. The linguistic distance has been computed using a δ value of 0.5. Column (1) shows the direct effect of ethnic group economic status on the likelihood of identifying more with one's ethnic group than with the nation. Column (2) adds an interaction term between economic status and linguistic distance, capturing how the effect of status varies with it. Column (3) replicates the specification in Column (2), but uses a normalized measure of linguistic distance within each country. Columns (4) through (6) mirror the specifications in Columns (1) through (3), respectively, with the addition of individual-level control variables.

Column (1) suggests that an increase in the economic status of an ethnic group is associated with a higher likelihood of identifying with the ethnic group. In Column (2), I observe that this effect is primarily driven by individuals who are linguistically distant from the rest of the country, as captured by the positive and significant coefficient of the interaction term, β_2. This intuition is confirmed in Column (3), where linguistic distance is standardized within each country, and both β_1 and β_2 are positive. In summary, improvements in the economic conditions of ethnic groups increase ethnic identification, but only among groups that are ethnically distinct. These results are consistent with the literature on secessionism, which finds that economic factors influence secessionist sentiment primarily among identity-based groups (Desmet et al., 2025; Gehring and Schneider, 2020).

The findings remain robust when individual-level controls are included. The reduction in the number of observations in these specifications is due to missing data on household occupation in Rounds 4 and 5 of the Afrobarometer survey.

The variable $Status_{ect}$ ranges from 0 to 51.89, while the linguistic distance measure ranges from 0 to 1 (see Table A2 in Appendix Section 4.3.4). This difference in scale may introduce multicollinearity when including both variables and their interaction in the regression. To address this, I rescale the variable $Status_{ect}$ as the ratio of a group's status to the maximum status observed for that group during the sample period. The results, reported in Table B1 in Appendix Section 4.A.1, are similar to those from the baseline specification.

Table 4.3. Economic Status and Linguistic Distance ($\delta = 0.5$): Social Identification

	(1)	(2)	(3)	(4)	(5)	(6)
Status	0.006**	-0.010	0.007***	0.011***	-0.008	0.012***
	(0.003)	(0.006)	(0.003)	(0.004)	(0.009)	(0.004)
Status x DistanceF		0.026***			0.031**	
		(0.010)			(0.014)	
Status x DistanceFstd			0.003**			0.003
			(0.002)			(0.002)
Indv Controls	No	No	No	Yes	Yes	Yes
Observations	101301	101301	101301	69082	69082	69082

Notes: Standard errors (in parentheses) are clustered by ethnic group-country-year. The dependent variable is a binary indicator equal to 1 if the individual identifies more with their ethnic group than with their nation. Status is defined as in Equation (1). Distance in columns (2) and (5) are defined as Equation (2), while in column (3) and (6) are normalized within the country. The δ used is equal to 0.5. Country-year, ethnicity, and cell fixed effects are included in all specifications. Columns (5)-(8) include individual controls: age, gender, rubal-rural residence and profession * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Results with δ = 0.05, and 0.0025. The main specification uses linguistic distance computed with $\delta = 0.5$ as this version shows the strongest effect in Table 4.2. On the other hand, Tables 4.4 and 4.5 present the results using alternative values of δ (0.05, and 0.0025, respectively). These results confirm the intuition suggested by Table 4.2: linguistic distances computed with lower values of δ have a weaker effect on social identification than those computed with higher values of δ.

Table 4.4. Economic Status and Linguistic Distance ($\delta = 0.05$): Social Identification

	(1)	(2)	(3)	(4)	(5)	(6)
Status	0.006**	-0.006	0.006**	0.011***	-0.000	0.011***
	(0.003)	(0.005)	(0.003)	(0.004)	(0.007)	(0.004)
Status x DistanceD		0.026**			0.027*	
		(0.010)			(0.015)	
Status x DistanceDstd			0.004***			0.004*
			(0.002)			(0.002)
Indv Controls	No	No	No	Yes	Yes	Yes
Observations	101301	101301	101301	69082	69082	69082

Notes: Standard errors (in parentheses) are clustered by ethnic group-country-year. The dependent variable is a binary indicator equal to 1 if the individual identifies more with their ethnic group than with their nation. Status is defined as in Equation (1) . Distance in columns (2) and (5) are defined as Equation (2), while in column (3) and (6) are normalized within the country. The δ used is equal to 0.05 (DistanceDesmet). Country-year, ethnicity, and cell fixed effects are included in all specifications. Columns (5)-(8) include individual controls: age, gender, rubal-rural residence and profession * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.5. Economic Status and Linguistic Distance ($\delta = 0.0025$): Social Identification

	(1)	(2)	(3)	(4)	(5)	(6)
Status	0.006**	-0.005	0.006**	0.011***	0.001	0.011***
	(0.003)	(0.005)	(0.003)	(0.004)	(0.007)	(0.004)
Status x DistanceG		0.026**			0.026*	
		(0.010)			(0.015)	
Status x DistanceGstd			0.004***			0.004**
			(0.002)			(0.002)
Indv Controls	No	No	No	Yes	Yes	Yes
Observations	101301	101301	101301	69082	69082	69082

Notes: Standard errors (in parentheses) are clustered by ethnic group-country-year. The dependent variable is a binary indicator equal to 1 if the individual identifies more with their ethnic group than with their nation. Status is defined as in Equation (1). Distance in columns (2) and (5) are defined as Equation (2), while in column (3) and (6) are normalized within the country. The δ used is equal to 0.0025. Country-year, ethnicity, and cell fixed effects are included in all specifications. Columns (5)-(8) include individual controls: age, gender, rubal-rural residence and profession * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.6 Robutness

Other ways to compute status. The political economy literature suggests that social identification is influenced not only by a group's absolute status, but also by its status relative to other groups. To capture this relative positioning, I normalize the economic status of each ethnic group using z-scores within their own country. Specifically, I standardize each group's status by subtracting the historical country-level mean and dividing by the country-level standard deviation, thereby measuring how far a group's status deviates from the national average over time. Results are reported in Table 4.6. Although the estimates are less statistically significant than the baseline, they suggest that ethnic groups with relatively high status are more likely to identify with their ethnic group, and that this effect is primarily driven by groups that are more ethnically distant from the rest of their country.

Table 4.6. Standarization of Economic Status and Linguistic Distance: Social Identification

	(1)	(2)	(3)	(4)	(5)	(6)
Status	0.013	-0.024	0.029**	0.024**	-0.027	0.039**
	(0.009)	(0.019)	(0.013)	(0.012)	(0.028)	(0.016)
Status x DistanceF		0.111*			0.161*	
		(0.059)			(0.087)	
Status x DistanceFstd			0.014*			0.013
			(0.007)			(0.009)
Indv Controls	No	No	No	Yes	Yes	Yes
Observations	101301	101301	101301	69082	69082	69082

Notes: Standard errors (in parentheses) are clustered by ethnic group-country-year. The dependent variable is a binary indicator equal to 1 if the individual identifies more with their ethnic group than with their nation. Status is defined based on how far a group's status deviates from their long-term country average. Distance in columns (2) and (5) are defined as Equation (2), while in column (3) and (6) are normalized within the country. The δ used is equal to 0.05. Country-year, ethnicity, and cell fixed effects are included in all specifications. Columns (5)-(8) include individual controls: age, gender, rubal-rural residence and profession * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Highly agricultural-dependent groups. Since my measure of status is based on the agricultural value of the ethnic homeland, and not all ethnic groups rely equally on agriculture, it is reasonable to expect that groups more dependent on agriculture will be more sensitive to the shock. To explore this, I rely on the work of Giuliano and Nunn (2018b), who map ethnolinguistic groups from Ethnologue to the Ethnographic Atlas (EA) compiled by Murdock (1967a). The EA contains information on the cultural practices of ethnic groups, including their level of dependence on agriculture. Using this dataset, I restrict the sample to ethnic groups whose agricultural dependence is at least 50% and replicate the estimates from Table 4.3. The results are reported in Appendix Section 4.A.3 in Table B4. They are very similar to those in the baseline specification.

Alternative classifications of Crops. The FAO Food Balance Sheets also provide information on crop imports and exports. To address the possibility that the definition of cash crops may introduce bias, I construct two alternative measures of $Status_{ect}$, distinct from the baseline measure based on cash crops. In the first alternative, I focus on the five major export crops with available international price of each country, calculating $Status_{ect}$ based on these crops. In the second, I focus on "net export" crops, defined as crops for which a country's exports exceed its imports. The results are reported in Appendix Section 4.A.4, in Tables B5 and B6. The direction of the results is consistent with the baseline, although the estimates are slightly less statistically significant.

4.7 Discussion

My results suggest that linguistic distance—a proxy for cultural distance—shapes social identification. This interpretation aligns with research in social psychology on how individuals form group attachments. The Minimal Group Paradigm (MGP), introduced by Tajfel (1970), shows that people often rely on simple, easily observable cues to form group associations, using these cues as the basis for social identity. In this context, my findings suggest that language can serve as a salient marker of group identity. The more linguistically distinct an individual is from the broader community, the more likely they are to identify with their ethnic group, reinforcing group boundaries and deepening in-group identification.

The role of linguistic markers as signals of group membership has also been widely explored in sociolinguistics. Ethnolinguistic Identity Theory (Giles, 1977) argues that accents and dialects function as powerful social cues that facilitate group affiliation and identity formation. More recently, Kinzler et al. (2007) found that children use accent as a key factor in social categorization, showing a preference for individuals who speak with a familiar accent when choosing potential friends. These findings can be particularly relevant in highly heterogeneous societies, such as those in sub-Saharan Africa, where ethnicity serves as a major social category.

4.8 Conclusion

The starting point of this project is the conceptual framework developed by Shayo (2009), in which individuals choose social identities based on a trade-off between the benefits of associating with high-status groups and the cognitive or cultural costs of identifying with groups that are distant from who they are. Focusing on sub-Saharan Africa, this study provides empirical evidence on how ethnic group economic conditions and cultural distance shape individual identification with ethnic versus national identities.

The findings show that individuals who are more linguistically distant from the rest of their country are more likely to identify with their ethnic group. I interpret linguistic distance as a proxy for the cognitive cost of national identification: the greater the perceived cultural distance from the nation, the higher the cost of identifying with it, leading individuals to align more strongly with their ethnic group. I also find that individuals from ethnic groups with better economic conditions are more likely to identify with their group, suggesting that group status enhances the perceived benefits of group affiliation. Taken together, these results provide empirical support for the cost-benefit framework proposed by Shayo (2009), highlighting how both economic benefits and cultural distance shape patterns of social identification.

Appendix

4..1 Classification of Countries and cash Crops

Table A1. Cash crops by country

Country	Cash Crops
Botswana	apple, banana, barley, cocoa, coconut, coffee, cotton, groundnuts, maize, oats, olive, orange, palmoil, rapeseedoil, rice, sorghum, soybean, sugar, sunfloweroil, tea, tomato, wheat
Ghana	apple, banana, barley, cocoa, coconut, coffee, cotton, groundnuts, maize, oats, olive, orange, palmoil, rapeseedoil, rice, sorghum, soybean, sugar, sunfloweroil, tea, tomato, wheat
Lesotho	apple, banana, barley, cocoa, coffee, maize, oats, orange, rice, sorghum, sugar, tea, tomato, wheat
Malawi	apple, banana, barley, cocoa, coconut, coffee, cotton, groundnuts, maize, oats, olive, orange, palmoil, rapeseedoil, rice, sorghum, soybean, sugar, sunfloweroil, tea, tomato, wheat
Namibia	apple, banana, barley, cocoa, coconut, coffee, cotton, groundnuts, maize, oats, olive, orange, rapeseedoil, rice, sorghum, soybean, sugar, sunfloweroil, tea, tomato, wheat
Nigeria	apple, barley, cocoa, coconut, coffee, cotton, groundnuts, maize, oats, olive, orange, palmoil, rapeseedoil, rice, sorghum, soybean, sugar, sunfloweroil, tea, tomato, wheat
South Africa	apple, banana, barley, cocoa, coconut, coffee, cotton, groundnuts, maize, oats, olive, orange, palmoil, rapeseedoil, rice, sorghum, soybean, sugar, sunfloweroil, tea, tomato, wheat
Tanzania	apple, banana, barley, cocoa, coconut, coffee, cotton, groundnuts, maize, oats, olive, orange, palmoil, rapeseedoil, rice, sorghum, soybean, sugar, sunfloweroil, tea, tomato, wheat
Uganda	apple, banana, barley, cocoa, coconut, coffee, cotton, groundnuts, maize, oats, olive, orange, palmoil, rapeseedoil, rice, sorghum, soybean, sugar, sunfloweroil, tea, tomato, wheat
Zambia	apple, banana, barley, cocoa, coconut, coffee, cotton, groundnuts, maize, oats, olive, orange, palmoil, rapeseedoil, rice, sorghum, soybean, sugar, sunfloweroil, tea, tomato, wheat
Zimbabwe	apple, banana, barley, cocoa, coconut, coffee, cotton, groundnuts, maize, oats, olive, orange, palmoil, rapeseedoil, rice, sorghum, soybean, sugar, sunfloweroil, tea, tomato, wheat

4..2 Summary Statistics

Table A2. Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Identity	0.18	0.38	0	1	101358
DistanceFearon	0.37	0.25	0.03	0.99	101358
DistanceDesmet	0.23	0.25	0	0.98	101358
DistanceGomes	0.22	0.26	0	0.97	101358
Status	4.63	6.12	0	51.89	101358
Gender	1.5	0.5	1	2	101351
Rural	1.78	0.42	1	3	101358
Age	36.98	15.04	18	115	100662

Notes: This table reports summary statistics for the main variables used in the analysis, as well as for the individual-level controls. Identity is a binary variable equal 1 if the individual identifies more with their ethnic group than with the nation. DistanceFearon, DistanceDesmet, and DistanceGomes refer to the linguistic distance defined in Equation 2, calculated using different values of δ: 0.5, 0.05, and 0.0025, respectively. Status is defined as Equation 1. Gender equals 1 if the individual is female. Rural takes value 1 if the individual resides in a rural area.

Table A3. Summary Statistics of Identity by Status Quartiles

Status Quartile	Mean Identity	Standard Deviation
Low	0.167	0.373
Mid-Low	0.180	0.383
Mid-High	0.182	0.386
High	0.185	0.388
Total	0.178	0.383

Notes: This table reports the mean and standard deviation of the binary dependent variable Identity, which takes value 1 if individual identity more with their ethnic group than with their nation, by quartiles of Status defined as Equation 4.1. The quartiles are based on the distribution of ethnic group-level status.

Table A4. Distribution of Identity

Identity	Frequency	Percent	Cumulative
0	83,317	82.20%	82.20%
1	18,041	17.80%	100.00%
Total	101,358	100.00%	

Notes: This table displays the frequency distribution of the binary dependent variable Identity, which is equal to 1 if the individual identifies more with their ethnic group than with the nation

Table A5. Detailed Summary Statistics for variables $Status_{ect}$

Statistic	Value
Observations	107,518
Mean	4.474
Standard Deviation	6.042
Variance	36.506
Skewness	1.919
Kurtosis	7.241
Percentiles	
1st Percentile	0.006
5th Percentile	0.019
10th Percentile	0.061
25th Percentile	0.382
50th Percentile (Median)	1.835
75th Percentile	5.359
90th Percentile	14.830
95th Percentile	18.237
99th Percentile	21.847
Maximum	51.881

Notes: This table reports detailed descriptive statistics for the variable Status as defined in Equation 1

Table A6. Between and Within variation: Status

Variable	Mean	Std. Dev.	Min	Max	Observations
Overall	5.304088	7.66481	0.0002774	51.88811	$N = 709$
Between	7.301646	0.0005443	39.99432		$n = 95$
Within	2.097536	-14.3303	17.19788		$T\text{-}bar = 7.46316$

Notes: This table reports the within and between varition of my variable Status from the ethnic groups presented in my sample.

Table A7. Cross-correlation table

Variables	Identity	DistanceFearon	DistanceDesmet	DistanceGomes	Status
Identity	1.00				
DistanceFearon	0.08	1.00			
DistanceDesmet	0.06	0.94	1.00		
DistanceGomes	0.06	0.92	1.00	1.00	
Status	-0.02	0.26	0.16	0.14	1.00

Notes: This table reports the cross-correlations between the main variables used in the analysis and the binary dependent variable Identity, which takes value 1 if individual identity more with their ethnic group than their country.

Table A8. Cross-correlation table: Individual variables

Variables	Identity	Gender	Rural	Age
Identity	1.00			
Gender	0.02	1.00		
Rural	-0.01	-0.00	1.00	
Age	-0.01	-0.07	-0.04	1.00

Notes: This table reports the cross-correlations between the main individual control variables and the binary dependent variable Identity, which takes value 1 if individual identity more with their ethnic group than their country.

4.A Appendix: Robustness

4.A.1 Other way of computing status

Table B1. Ratio of Economic Status and Linguistic Distance ($\delta = 0.5$): Social Identification

	(1)	(2)	(3)	(4)	(5)	(6)
StatusRatio	0.012**	-0.021	0.014***	0.023***	-0.017	0.025***
	(0.006)	(0.013)	(0.005)	(0.008)	(0.019)	(0.008)
StatusRatio x DistanceF		0.054***			0.065**	
		(0.020)			(0.029)	
StatusRatio x DistanceFstd			0.007**			0.006
			(0.003)			(0.004)
Indv Controls	No	No	No	Yes	Yes	Yes
Observations	101301	101301	101301	69082	69082	69082

Notes: Standard errors (in parentheses) are clustered by ethnic group-country-year. The dependent variable is a binary indicator equal to 1 if the individual identifies more with their ethnic group than with their nation. Status is defined as the ratio of a group's status to the maximum status observed for that group during the sample period. Distance in columns (2) and (5) are defined as Equation (2), while in column (3) and (6) are normalized within the country. The δ used is equal to 0.5. Country-year, ethnicity, and cell fixed effects are included in all specifications. Columns (5)-(8) include individual controls: age, gender, rubal-rural residence and profession * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.A.2 Clustering at ethnic group-country level

Table B2. Cultural Distance and Individual Identification

	(1)	(2)	(3)	(4)	(5)	(6)
DistanceGomes	0.055**			0.053**		
	(0.023)			(0.027)		
DistanceDesmet		0.061**			0.057**	
		(0.023)			(0.028)	
DistanceFearon			0.101***			0.084**
			(0.032)			(0.035)
Observations	101301	101301	101301	69082	69082	69082

Notes: Standard errors (in parentheses) are clustered by ethnic group-country. The dependent variable is a binary indicator equal to 1 if the individual identifies more with their ethnic group than with their nation. Distance is defined as in Equation 2 at varying levels of aggregation: DistanceFearon corresponds to $\delta = 0.5$, DistanceDesmet to $\delta = 0.05$, and DistanceGomes to $\delta = 0.0025$. Country-year and cell fixed effects are included in all spefications. Columns (4)-(6) include specifications individual controls * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B3. Economic Status and Linguistic Distance ($\delta = 0.5$): Social Identification

	(1)	(2)	(3)	(4)	(5)	(6)
Status	0.006	-0.010	0.007**	0.011**	-0.008	0.012**
	(0.004)	(0.007)	(0.003)	(0.005)	(0.010)	(0.005)
Status x DistanceF		0.026**			0.031**	
		(0.010)			(0.015)	
Status x DistanceFstd			0.003*			0.003
			(0.002)			(0.002)
Indv Controls	No	No	No	Yes	Yes	Yes
Observations	101301	101301	101301	69082	69082	69082

Notes: Standard errors (in parentheses) are clustered by ethnic group-country level. The dependent variable is a binary indicator equal to 1 if the individual identifies more with their ethnic group than with their nation. Status is defined as in Equation (1). Distance in columns (2) and (5) are defined as Equation (2), while in column (3) and (6) are normalized within the country. The δ used is equal to 0.5. Country-year, ethnicity and cell fixed effects are included in all speciations. Columns (5)-(8) include individual controls: age, gender, urban-rural residence and profession * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.A.3 Highly Dependent Groups

Table B4. Economic Status and Linguistic Distance ($\delta = 0.5$): Social Identification

	(1)	(2)	(3)	(4)	(5)	(6)
Status	0.006**	-0.010	0.007***	0.012***	-0.006	0.013***
	(0.003)	(0.007)	(0.003)	(0.004)	(0.009)	(0.004)
Status x DistanceF		0.025**			0.029*	
		(0.010)			(0.015)	
Status x DistanceFstd			0.003*			0.003
			(0.002)			(0.002)
Observations	93620	93620	93620	63882	63882	63882

Notes: Standard errors (in parentheses) are clustered by ethnic group-country year. The dependent variable is a binary indicator equal to 1 if the individual identifies more with their ethnic group than with their nation. Status is defined as in Equation (1). Distance in columns (2) and (5) are defined as Equation (2), while in column (3) and (6) are normalized within the country. The δ used is equal to 0.5. Country-year, ethnicity and cell fixed effects are included in all speciations. Columns (5)-(8) include individual controls: age, gender, urban-rural residence and profession * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.A.4 Other classification of crops

Table B5. Economic Status (Major Crops) and Linguistic Distance: Social Identification

	(1)	(2)	(3)	(4)	(5)	(6)
Status	0.003^{*}	-0.004	0.004^{**}	0.005	-0.003	0.006^{*}
	(0.002)	(0.003)	(0.002)	(0.003)	(0.005)	(0.004)
Status x DistanceF		0.018^{***}			0.021^{*}	
		(0.006)			(0.012)	
Status x DistanceFstd			0.003^{*}			0.002
			(0.002)			(0.002)
Observations	101301	101301	101301	69082	69082	69082

Notes: Standard errors (in parentheses) are clustered by ethnic group-country level. The dependent variable is a binary indicator equal to 1 if the individual identifies more with their ethnic group than with their nation. Status is defined as in Equation (1), but calculated using the five major export crops for which international price data are available. Distance in columns (2) and (5) are defined as Equation (2), while in column (3) and (6) are normalized within the country. The δ used is equal to 0.5. Country-year, ethnicity and cell fixed effects are included in all spefications. Columns (5)-(8) include individual controls: age, gender, urban-rural residence and profession * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B6. Economic Status (Net Crops) and Linguistic Distance: Social Identification

	(1)	(2)	(3)	(4)	(5)	(6)
Status	0.004^{*}	-0.003	0.005^{**}	0.005	0.004	0.007
	(0.002)	(0.006)	(0.002)	(0.004)	(0.009)	(0.004)
Status x DistanceF		0.013			0.002	
		(0.011)			(0.018)	
Status x DistanceFstd			0.003^{*}			0.003
			(0.002)			(0.002)
Observations	101301	101301	101301	69082	69082	69082

Notes: Standard errors (in parentheses) are clustered by ethnic group-country level. The dependent variable is a binary indicator equal to 1 if the individual identifies more with their ethnic group than with their nation. Status is defined as in Equation (1), but calculated using crops for which a country's exports exceed its imports. Distance in columns (2) and (5) are defined as Equation (2), while in column (3) and (6) are normalized within the country. The δ used is equal to 0.5. Country-year, ethnicity and cell fixed effects are included in all spefications. Columns (5)-(8) include individual controls: age, gender, urban-rural residence and profession * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Bibliography

Ahlerup, Pelle, Thushyanthan Baskaran, and Arne Bigsten, "Regional development and national identity in sub-Saharan Africa," *Journal of Comparative Economics*, 2017, *45* (3), 622–643.

Aker, Jenny C., Michael W. Klein, Stephen A. O'Connell, and Muzhe Yang, "Borders, Ethnicity and Trade," *Journal of Development Economics*, 2014, *107* (C), 1–16.

Akerlof, George A and Rachel E Kranton, "Economics and identity," *The quarterly journal of economics*, 2000, *115* (3), 715–753.

— **and** — , *Identity economics: How our identities shape our work, wages, and well-being*, Princeton University Press, 2010.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg, "Fractionalization," *Journal of Economic Growth*, 2003, *8, no. 2, June*, 155–194.

Alesina, Alberto, Stelios Michalopoulos, and Elias Papaioannou, "Ethnic Inequality," *Journal of Political Economy*, 2016, *124* (2), 428–488.

Ansolabehere, Stephen and M Socorro Puy, "Identity voting," *Public Choice*, 2016, *169* (1), 77–95.

Arbatlı, Cemal Eren, Quamrul H Ashraf, Oded Galor, and Marc Klemp, "Diversity and Conflict," *Econometrica*, 2020, *88* (2), 727–797.

Armand, Alex, Paul Atwell, and Joseph Gomes, "The Reach of Radio: Ending Civil Conflict through Rebel Demobilization," *American Economic Review*, 2020, *110* (5), 1–36.

Autor, David H, David Dorn, and Gordon H Hanson, "The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade," *Annual Review of Economics*, 2016, *8*, 205–240.

_ , _ , _ , and Jae Song, "Trade Adjustment: Worker-Level Evidence," *Quarterly Journal of Economics*, 2014, *129* (4), 1799–1860.

Banchirigah, Sadia Mohammed and Gavin Hilson, "De-agrarianization, re-agrarianization and local economic development: Re-orientating livelihoods in African artisanal mining communities," *Policy Sciences*, 2010, *43* (2), 157–180.

Barbieri, Katherine and Rafael Reuveny, "Economic Globalization and Civil War," *Journal of Politics*, 2005, *67* (4), 1228–1247.

Bazillier, Remi and Victoire Girard, "The gold digger and the machine. Evidence on the distributive effect of the artisanal and industrial gold rushes in Burkina Faso," *Journal of Development Economics*, 2020, *143*, 102411.

Bazzi, Samuel and Christopher Blattman, "Economic Shocks and Conflict: Evidence from Commodity Prices," *American Economic Journal: Macroeconomics*, 2014, *6* (4), 1–38.

Berman, Nicolas and Mathieu Couttenier, "External Shocks, Internal Shots: The Geography of Civil Conflicts," *Review of Economics and Statistics*, 2015, *97* (4), 758–776.

_ , _ , and Victoire Girard, "Mineral resources and the salience of ethnic identities," *The Economic Journal*, 2023, *133* (653), 1705–1737.

_ , _ , Dominic Rohner, and Mathias Thoenig, "This Mine is Mine! How Minerals Fuel Conflicts in Africa," *American Economic Review*, 2017, *107* (6), 1564–1610.

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

Blair, Graeme, Darin Christensen, and Aaron Rudkin, "Do Commodity Price Shocks Cause Armed Conflict? A Meta-Analysis of Natural Experiments," *American Political Science Review*, 2021, *115* (2), 709–716.

Bó, Ernesto Dal and Pedro Dal Bó, "Workers, Warriors, and Criminals: Social Conflict in General Equilibrium," *Journal of the European Economic Association*, 2011, *9* (4), 646–677.

Boken, Johannes, Lucie Gadenne, Tushar Nandi, and Marta Santamaria, "Community Networks and Trade," CEPR Discussion Paper 17787 2023.

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria et al., "The value of democracy: evidence from road building in Kenya," *The American Economic Review*, 2015, *105* (6), 1817–1851.

Burke, Marshall, Solomon M Hsiang, and Edward Miguel, "Climate and Conflict," *Annual Review of Economics*, 2015, *7* (1), 577–617.

Cervellati, Matteo, Elena Esposito, and Uwe Sunde, "Epidemic Shocks and Civil Violence: Evidence from Malaria Outbreaks in Africa," *Review of Economics and Statistics*, 2022, *104* (4), 780–796.

Christiaensen, Luc and Will Martin, "Agriculture, structural transformation and poverty reduction: Eight new insights," 2018.

Ciccone, Antonio, "Economic Shocks and Civil Conflict: A Comment," *American Economic Journal: Applied Economics*, 2011, *3* (4), 215–27.

Collier, P. and A. Hoeffler, "Greed and Grievance in Civil War," *Oxford Economic Papers*, 2004, *56(4)*, 563–596.

Conley, Timothy G, "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

David, H, David Dorn, and Gordon H Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–68.

Dell, Melissa, "Trafficking networks and the Mexican drug war," *American Economic Review*, 2015, *105* (6), 1738–1779.

Depetris-Chauvin, Emilio, Ruben Durante, and Filipe Campante, "Building nations through shared experiences: Evidence from African football," *American Economic Review*, 2020, *110* (5), 1572–1602.

Desmet, Klaus, Ignacio Ortu no Ortín, and I. Weber, "Linguistic Diversity and Redistribution," *Journal of European Economic Association*, 2009, *7(6)*, 1291–1318.

_ , Ignacio Ortuño Ortín, and Ömer Özak, "Is Secessionism Mostly About Income or Identity? A Global Analysis of 3,153 Subnational Regions," *The Economic Journal*, 2025, *135* (668), 1261–1299.

_ , Ignacio Ortuño-Ortín, and Romain Wacziarg, "The Political Economy of Linguistic Cleavages," *Journal of Development Economics*, 2012, *97* (2), 322–338.

_ , Joseph Flavian Gomes, and Ignacio Ortuño-Ortín, "The Geography of Linguistic Diversity and the Provision of Public Goods," *Journal of Development Economics*, 2020, *143*, 102384.

_ , Joseph Gomes, and Diego Malo Rico, "Ethnic Remoteness Reduces the Peace Dividend from Trade Access," *Journal Name*, 2024. Forthcoming.

_ , Shlomo Weber, and Ignacio Ortuño-Ortín, "Linguistic diversity and redistribution," *Journal of the European Economic Association*, 2009, *7* (6), 1291–1318.

Dix-Carneiro, Rafael, Rodrigo R Soares, and Gabriel Ulyssea, "Economic Shocks and Crime: Evidence from the Brazilian Trade Liberalization," *American Economic Journal: Applied Economics*, 2018, *10* (4), 158–95.

Dube, Oeindrila and Juan F Vargas, "Commodity Price Shocks and Civil Conflict: Evidence from Colombia," *Review of Economic Studies*, 2013, *80* (4), 1384–1421.

Dun & Bradstreet, "Dun & Bradstreet," https://www.dnb.com/ D&B. Accessed: 2023-11 - 2023-12.

Easterly, W. and R. Levine, "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 1997, *112, no.4 November*, 1203–1250.

Eifert, Benn, Edward Miguel, and Daniel N Posner, "Political competition and ethnic identification in Africa," *American journal of political science*, 2010, *54* (2), 494–510.

Enke, Benjamin, "Kinship, Cooperation, and the Evolution of Moral Systems," *Quarterly Journal of Economics*, 2019, *134* (2), 953–1019.

Esteban, Joan, Laura Mayoral, and Debraj Ray, "Ethnicity and Conflict: An Empirical Study," *American Economic Review*, 2012, *102, No.4*, 1310–1342.

_ , _ , and _ , "Ethnicity and Conflict: Theory and Facts," *Science*, 2012, *336, 858.*

Esteban, Joan Maria and Debraj Ray, "On the Measurement of Polarization," *Econometrica*, 1994, *62, No.4*, 819–851.

FAO and IIASA, *Global Agro-Ecological Zones (GAEZ v3.0)* 2012.

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

Fearon, James D., "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth*, 2003, *8* (2), 195–222.

Fearon, James D, "Ethnic and cultural diversity by country," *Journal of economic growth*, 2003, *8*, 195–222.

Fenske, James and Namrata Kala, "Linguistic Distance and Market Integration in India," *Journal of Economic History*, 2021, *81* (1), 1–39.

Fernandes, Ana M, Alejandro Forero, Hibret Maemir, and Aaditya Mattoo, "Are Trade Preferences a Panacea? The Export Impact of the African Growth and Opportunity Act," *World Development*, 2023, *162*, 106114.

Fetzer, Thiemo, "Can Workfare Programs Moderate Conflict? Evidence from India," *Journal of the European Economic Association*, 2020, *18* (6), 3337–3375.

Franck, Raphael and Ilia Rainer, "Does the leader's ethnicity matter? Ethnic favoritism, education, and health in sub-Saharan Africa," *American Political Science Review*, 2012, *106* (2), 294–325.

Francois, Patrick, Ilia Rainer, and Francesco Trebbi, "How is Power Shared in Africa?," *Econometrica*, 2015, *83* (2), 465–503.

Frazer, Garth and Johannes Van Biesebroeck, "Trade Growth under the African Growth and Opportunity Act," *Review of Economics and Statistics*, 2010, *92* (1), 128–144.

Funk, Chris C, Pete J Peterson, Martin F Landsfeld, Diego H Pedreros, James P Verdin, James D Rowland, Bo E Romero, Gregory J Husak, Joel C Michaelsen, Andrew P Verdin et al., "A Quasi-Global Precipitation Time Series for Drought Monitoring," *US Geological Survey Data Series*, 2014, *832* (4), 1–12.

Gaulier, Guillaume and Soledad Zignago, "BACI: International Trade Database at the Product-Level. The 1994-2007 Version," Working Papers 2010-23, CEPII 2010.

Gehring, Kai and Stephan A Schneider, "Regional resources and democratic secessionism," *Journal of Public Economics*, 2020, *181*, 104073. *Geofabrik Download Server*

Geofabrik Download Server.

Giles, Howard, *"Social psychology and applied linguistics: Towards an integrative approach," ITL-International Journal of Applied Linguistics, 1977, 35 (1), 27–42.*

Gillard, Tyler, *"OECD Due Diligence Guidance for Responsible Supply Chains of Minerals from Conflict-Affected and High-Risk Areas," 2015.*

Girard, Victoire, Teresa Molina-Millán, Guillaume Vic et al., *"Artisanal mining in Africa," in "Working Paper Series. No. 2201" 2022.*

Giuliano, Paola and Nathan Nunn, *"Ancestral Characteristics of Modern Populations," Economic History of Developing Regions, 2018, 33 (1), 1–17.*

_ and _ , *"Ancestral characteristics of modern populations," Economic History of Developing Regions, 2018, 33 (1), 1–17.*

Goldberg, Pinelopi Koujianou and Nina Pavcnik, *"Distributional Effects of Globalization in Developing Countries," Journal of Economic Literature, 2007, 45 (1), 39–82.*

Gomes, Joseph Flavian, *"The Health Costs of Ethnic Distance: Evidence from Sub-Saharan Africa," Journal of Economic Growth, 2020, 25, 195–226.*

Green, Elliott, *"Ethnicity, national identity and the state: Evidence from sub-Saharan Africa," British Journal of Political Science, 2020, 50 (2), 757–779.*

Greenberg, Joseph H, *"The Measurement of Linguistic Diversity," Language, 1956, 32 (1), 109–115.*

Grossman, Gene M and Elhanan Helpman, *"Identity politics and trade policy," The Review of Economic Studies, 2021, 88 (3), 1101–1126.*

Guarnieri, Eleonora, *"Cultural Distance and Ethnic Civil Conflict," 2024.*

_ **and Ana Tur-Prats**, *"Cultural distance and conflict-related sexual violence," The Quarterly Journal of Economics, 2023, 138 (3), 1817–1861.*

Guriev, Sergei and Elias Papaioannou, *"The political economy of populism," Journal of Economic literature, 2022, 60 (3), 753–832.*

Hansen, Matthew C, Peter V Potapov, Rebecca Moore, Matt Hancher, Svetlana A Turubanova, Alexandra Tyukavina, David Thau, Stephen V Stehman, Scott J Goetz, Thomas R Loveland et al., *"High-resolution global maps of 21st-century forest cover change," science, 2013, 342 (6160), 850–853.*

HarvestChoice, *"Travel Time to Nearest Port (Hours, 2000)," Technical Report, International Food Policy Research Institute, Washington, D.C. 2015.*

Henderson, J Vernon, Adam Storeygard, and David N Weil, *"Measuring Economic Growth from Outer Space," American Economic Review, 2012, 102 (2), 994–1028.*

Hidalgo, F Daniel, Suresh Naidu, Simeon Nichter, and Neal Richardson, *"Economic Determinants of Land Invasions," Review of Economics and Statistics, 2010, 92 (3), 505–523.*

Hilson, Gavin, Abigail Hilson, Roy Maconachie, James McQuilken, and Halima Goumandakoye, *"Artisanal and small-scale mining (ASM) in sub-Saharan Africa: Re-conceptualizing formalization and 'illegal'activity," Geoforum, 2017, 83, 80–90.*

Hsiang, Solomon M, *"Temperatures and Cyclones Strongly Associated with Economic Production in the Caribbean and Central America," Proceedings of the National Academy of Sciences, 2010, 107 (35), 15367–15372.*

Hunter, Marcena, *"Pulling at golden webs," Combating criminal consortia in the Africa n artisanal and small-scale gold mining and trade sector. ENACT, 2019.*
Illegal Gold Mining in Central Africa

Illegal Gold Mining in Central Africa.

Isphording, Ingo E, "Disadvantages of Linguistic Origin—Evidence from Immigrant Literacy Scores," *Economics Letters*, 2014, *123* (2), 236–239.

Isphording, Ingo Eduard and Sebastian Otten, "The Costs of Babylon—Linguistic Distance in Applied Economics," *Review of International Economics*, 2013, *21* (2), 354–369.

Jacobson, Louis S, Robert J LaLonde, and Daniel G Sullivan, "Earnings losses of displaced workers," *The American economic review*, 1993, pp. 685–709.

Jha, Saumitra, "Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia," *American Political Science Review*, 2013, *107 (4)*, 806–832.

Justino, Patricia, Santiago Tobón, Martín Vanegas-Arias, and Juan F Vargas, *Concealed costs: illicit economies and the erosion of the local tax base in Colombia* number 2024/62, WIDER Working Paper, 2024.

Karakas, Leyla D and Devashish Mitra, "Immigration policy in the presence of identity politics," Technical Report, Working Paper 2019.

Kinzler, Katherine D, Emmanuel Dupoux, and Elizabeth S Spelke, "The native language of social cognition," *Proceedings of the National Academy of Sciences*, 2007, *104* (30), 12577–12580.

Kiszewski, Anthony, Andrew Mellinger, Andrew Spielman, Pia Malaney, Sonia Ehrlich Sachs, and Jeffrey Sachs, "A Global Index Representing the Stability of Malaria Transmission," *American Journal of Tropical Medicine and Hygiene*, 2004, *70* (5), 486–498.

Krugman, Paul R, *Geography and Trade*, MIT Press, 1991.

Laitin, David D and Rajesh Ramachandran, "Language Policy and Human Development," *American Political Science Review*, 2016, *110* (3), 457–480.

Lebrand, Mathilde Sylvie Maria, Hannes Felix Mueller, Peer Schouten, and Jevgenijs Steinbuks, "Road Investment and Violence in DRC: Perishable Peace Dividends," *Available at SSRN 5273313*, 2025.

Lewis, M Paul, Gary F Simons, Charles D Fennig et al., *Ethnologue: Languages of the World*, Vol. 17, SIL international Dallas, TX, 2014.

Luca, Giacomo De, Roland Hodler, Paul A Raschky, and Michele Valsecchi, "Ethnic favoritism: An axiom of politics?," *Journal of Development Economics*, 2018, *132*, 115–129.

Lujala, Päivi, Jan Ketil Rød, and Nadja Thieme, "Fighting over Oil: Introducing a New Dataset," *Conflict Management and Peace Science*, 2007, *24* (3), 239–256.

Martin, Philippe, Thierry Mayer, and Mathias Thoenig, "Civil Wars and International Trade," *Journal of the European Economic Association*, 2008, *6* (2-3), 541–550.

_ , _ , and _ , "Make Trade not War?," *Review of Economic Studies*, 2008, *75* (3), 865–900.

_ , _ , and _ , "The Geography of Conflicts and Regional Trade Agreements," *American Economic Journal: Macroeconomics*, 2012, *4* (4), 1–35.

Matthysen, Ken, Thomas Muller, Jan Leysen, and Ntakobajira Zacharie Bulakali, "Analysis of the Interactive Map of Artisanal Mining Areas in Eastern Democratic Republic of Congo (2023 Update)," *IPIS, upd Dec*, 2023, p. 46.

McCaig, Brian, "Exporting out of Poverty: Provincial Poverty in Vietnam and U.S. Market Access," *Journal of International Economics*, 2011, *85* (1), 102–113.

McCord, Gordon C and Jesse K Anttila-Hughes, "A Malaria Ecology Index Predicted Spatial and Temporal Variation of Malaria Burden and Efficacy of Antimalarial Interventions Based on African Serological Data," *American Journal of Tropical Medicine and Hygiene*, 2017, *96* (3), 616–623.

McGuirk, Eoin and Marshall Burke, "The Economic Origins of Conflict in Africa," *Journal of Political Economy*, 2020, *128* (10), 3940–3997.

Melitz, Jacques and Farid Toubal, "Native Language, Spoken Language, Translation and Trade," *Journal of International Economics*, 2014, *93* (2), 351–363.

Meyer, Bruce D. and James X. Sullivan, "Measuring the Well-Being of the Poor Using Income and Consumption," *Journal of Human Resources*, 2003, *38*, 1180–1220.

Michalopoulos, Stelios and Elias Papaioannou, "Pre-colonial ethnic institutions and contemporary African development," *Econometrica*, 2013, *81* (1), 113–152.

_ and _, "Spatial Patterns of Development: A Meso Approach," *Annual Review of Economics*, 2018, *10*, 383–410.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti, "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," *Journal of Political Economy*, 2004, *112* (4), 725–753.

Mitra, Anirban and Debraj Ray, "Implications of an Economic Theory of Conflict: Hindu-Muslim Violence in India," *Journal of Political Economy*, 2014, *122* (4), 719–765.

Montalvo, José G and Marta Reynal-Querol, "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review*, 2005, *95* (3), 796–816.

Moscona, Jacob, Nathan Nunn, and James A Robinson, "Segmentary Lineage Organization and Conflict in Sub-Saharan Africa," *Econometrica*, 2020, *88* (5), 1999–2036.

Müller, Lars, "The Force Intervention Brigade—United Nations forces beyond the fine line between peacekeeping and peace enforcement," *Journal of Conflict and Security Law*, 2015, *20* (3), 359–380.

Muller, T, "IPIS open data dashboard on artisanal and small-scale mining in eastern DRC," 2020.

Muñoz-Sabater, Joaquín, Emanuel Dutra, Anna Agustí-Panareda, Clément Albergel, Gabriele Arduini, Gianpaolo Balsamo, Souhail Boussetta, Margarita Choulga, Shaun Harrigan, Hans Hersbach et al., "ERA5-Land: A State-of-the-Art Global Reanalysis Dataset for Land Applications," *Earth System Science Data Discussions*, 2021, pp. 1–50.

Murdock, George P., "Ethnographic Atlas," *Ethnology*, 1967, *6* (2), 109–236.

Murdock, George Peter, "Ethnographic Atlas: A Summary," *Ethnology*, 1967, *6* (2), 109–236.

National Oceanic and Atmospheric Administration, "DMSP-OLS Nighttime Lights Time Series Version 4," National Geophysical Data Center, United States Department of Commerce. https://www.ngdc.noaa.gov 2014.

Nechaev, Dmitry, Mikhail Zhizhin, Alexey Poyda, Tilottama Ghosh, Feng-Chi Hsu, and Christopher Elvidge, "Cross-Sensor Nighttime Lights Image Calibration for DMSP/OLS and SNPP/VIIRS with Residual U-Net," *Remote Sensing*, 2021, *13* (24), 5026.

Nunn, Nathan and Nancy Qian, "The Potato's Contribution to Population and Urbanization: Evidence from a Historical Experiment," *The quarterly journal of economics*, 2011, *126* (2), 593–650.

Ober, Joyce A, "Mineral commodity summaries 2018," *USGS Report*, 2018, p. 2.

OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org." Published: https://www.openstreetmap.org.

Pereira, Brais Álvarez, Martín Portos, and John Vourdas, "Waving goodbye? The determinants of autonomism and secessionism in Western Europe," *Regional Studies*, 2018, *52* (2), 197–211.

Pierce, Justin R. and Peter K. Schott, "A Concordance between Ten-Digit US Harmonized System Codes and SIC/NAICS Product Classes and Industries," *Journal of Economic and Social Measurement*, 2012, *37* (1-2), 61–96.

Raleigh, Clionadh and Caitriona Dowd, "Armed conflict location and event data project (ACLED) codebook," *Find this resource*, 2015.

_ , **Andrew Linke, Håvard Hegre, and Joakim Karlsen**, "Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature," *Journal of Peace Research*, 2010, *47* (5), 651–660.

_ , **Roudabeh Kishi, and Andrew Linke**, "Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices," *Humanities and Social Sciences Communications*, 2023, *10* (1), 1–17.

Ramankutty, Navin, Amato T Evan, Chad Monfreda, and Jonathan A Foley, "Farming the Planet: 1. Geographic Distribution of Global Agricultural Lands in the Year 2000," *Global Biogeochemical Cycles*, 2008, *22* (1).

Rohner, Dominic, Mathias Thoenig, and Fabrizio Zilibotti, "Seeds of distrust: Conflict in Uganda," *Journal of Economic Growth*, 2013, *18*, 217–252.

Ross, Michael L, "What do we know about natural resources and civil war?," *Journal of peace research*, 2004, *41* (3), 337–356.

Savoia, Antonio and Kunal Sen, "The political economy of the resource curse: A development perspective," *Annual Review of Resource Economics*, 2021, *13* (1), 203–223.

Sebastian, Kate, *Atlas of African Agriculture Research and Development: Revealing Agriculture's Place in Africa*, International Food Policy Research Institute, 2014.

Shastry, Gauri Kartini, "Human Capital Response to Globalization Education and Information Technology in India," *Journal of Human Resources*, 2012, *47* (2), 287–330.

Shayo, Moses, "A model of social identity with an application to political economy: Nation, class, and redistribution," *American Political science review*, 2009, *103* (2), 147–174.

Sierra, Raúl Sánchez De La, "On the origins of the state: Stationary bandits and taxation in eastern congo," *Journal of Political Economy*, 2020, *128* (1), 000–000.

Simple Maps, "World City Database," https://simplemaps.com/data/world-cities 2023. Accessed: 2023-12-14.

Spolaore, Enrico and Romain Wacziarg, "The Diffusion of Development," *Quarterly Journal of Economics*, 2009, *124* (2), 469–529.

Stoop, Nik, Marijke Verpoorten, and Peter Van der Windt, "More legislation, more violence? The impact of Dodd-Frank in the DRC," *PloS one*, 2018, *13* (8), e0201783.

Sundberg, Ralph and Erik Melander, "Introducing the UCDP Georeferenced Event Dataset," *Journal of Peace Research*, 2013, *50* (4), 523–532.

Tajfel, Henri, "Experiments in intergroup discrimination," *Scientific american*, 1970, *223* (5), 96–103.

_ , "Social identity and intergroup behaviour," *Social science information*, 1974, *13* (2), 65–93.

_ , "An integrative theory of intergroup conflict," *The social psychology of intergroup relations/Brooks/Cole*, 1979.

_ , "Human groups and social categories: Studies in social psychology," *Cup Archive*, 1981.

Titeca, Kristof and Giovanni Salvaggio, "Understanding Uganda's (ambiguous) actions in Eastern DRC: military interventions to protect roads and trade?," 2025.

Topalova, Petia, "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India," *American Economic Journal: Applied Economics*, 2010, *2* (4), 1–41.

Turner, J.C., *Rediscovering the Social Group: A Self-categorization Theory*, Basil Blackwell, 1987.

Zabyelina, Yuliya and Lilla Heins, "All that glitters: Money laundering through precious metals and minerals," in "Illegal mining: Organized crime, corruption, and ecocide in a resource-scarce world," Springer, 2020, pp. 439–465.