Essays on Gender-Based Violence and Ethnic Conflict

Eleonora Guarnieri

Dissertation

Department of Economics
LMU Munich

2021

Essays on Gender-Based Violence and Ethnic Conflict

Inaugural-Dissertation

zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Volkswirtschaftlichen Fakultät
der Ludwig-Maximilians-Universität München

2021

vorgelegt von

Eleonora Guarnieri

Referent:	Prof. Helmut Rainer, Ph.D.
Korreferent:	Prof. Davide Cantoni, Ph.D.
Mündliche Prüfung:	29. Juni 2021
Berichterstatter:	Prof. Helmut Rainer, Ph.D. Prof. Davide Cantoni, Ph.D. Prof. Ana Tur-Prats, Ph.D.
Promotionsabschlussberatung:	14. Juli 2021

Acknowledgements

This thesis is the culmination of an exciting, life-changing, and at times challenging journey. Many people helped me along the way, and I am profoundly grateful to each and every one of them.

Helmut Rainer, my advisor, made this journey possible in the first place. I am thankful to him not only for the opportunity to pursue a Ph.D., but also for all the encouragement and advice he offered in these past years, especially during the most difficult times. I further thank him for being my coauthor in the first chapter of this thesis: working with him was for me an invaluable learning experience. I owe a lot of gratitude to Davide Cantoni. He tirelessly helped me improve my work, and patiently guided me through the impervious routes of the academic job market. His support went far beyond the duties of a second advisor and placement coordinator. Meeting Ana Tur-Prats was a fortunate turnaround in my Ph.D. journey. She has been my role model, an inspiring coauthor, and a wise mentor. Our countless conversations fueled my enthusiasm and made me a better researcher. I deeply thank her, as writing this dissertation would not have been possible without her at my side.

During my Ph.D., I received generous funding from the ifo Institute and the Leibniz Association. I am grateful for the many opportunities to travel, present my research, and interact with innumerable inspiring people. One of them is Graziella Bertocchi, whom I sincerely thank. My research benefited tremendously from the time I spent in Boston. I am indebted to Seema Jayachandran and the J-PAL Gender Sector for allowing me to work with them, and to Anukriti for hosting me at Boston College.

During all these years, I was blessed with the support of many colleagues at the ifo Institute. I owe a particular thank you to my fellow Ph.D. students and friends Anna, Fabian, Marc, Pablo, Patrick, and Victoria. Without them, obstacles would have felt insurmountable and my long days at ifo would not have been as fun. I am infinitely grateful to Clara, who brought a ray of sunshine to my darkest days and always helped me put things into perspective. I am thankful to all my other friends, old and new. Kristin brightened up my free time in Munich and never forgot to check on me during the lonely months of the pandemic. Alessandra, Federica, Raffaele, Veronica, and Vittoria, my lifelong friends, have always been there despite the distance.

My family's love, support, and sacrifices are the foundations of who I am today. Giovanni, my brother, taught me to stay good-hearted, no matter what life puts me through. Onorina, my mom, taught me to strive for independence and to be ambitious, curious, and hard-working. I am indebted to her more than she will ever imagine. Luigi, my dad, fills my days with laughter, joy, and wisdom. He encouraged me to discover the world and never stops reminding me that, when things get tough, I must be *"immobile come lo scoglio sul lido del mare: i flutti lo flagellano, ma non lo crollano."* He is my greatest source of inspiration and the person I aspire to be. *Mamma, papà e Giovanni: grazie di cuore.*

Finally, my biggest *grazie* goes to Sebastian. He is my rock, and his love is what keeps me going.

Contents

List of Figures

List of Tables

Preface

"*I lived my childhood as a village girl in Kojo [...]. I knew nothing about the conflicts and killings that took place in our world every day. I did not know that human beings could perpetrate such hideous crimes against each other.*"
—Nadia Murad, The Nobel Peace Prize Lecture 2018

"*The macabre violence knew no limit. Sadly, this violence has never stopped.*"
—Dr. Denis Mukwege, The Nobel Peace Prize Lecture 2018

The civil war in the eastern Democratic Republic of Congo has killed 5.4 million people since 1996. This is believed to be the deadliest conflict since World War II. Dr. Denis Mukwege, a Congolese gynecologist, has treated thousands of women raped by armed rebels. In 2014, ISIS abducted and forced into sexual slavery thousands of Yazidi women and girls in northern Iraq, including Nadia Murad. Worldwide, domestic abuse kills 30,000 people a year (United Nations Office on Drugs and Crime [2018]). Violence—from the one perpetrated within the home to armed conflict—constitutes a fundamental human rights violation and imposes enormous economic costs on individuals, families, and societies. In 2019 alone, conservative estimates suggest that the global cost of violence and conflict amounted to 10.5 percent of the global GDP (Institute for Economics & Peace [2020]).

While the consequences of violence are well-documented, understanding *why* individuals and groups act violently remains an important object of study across the social sciences. Scholars in psychology, sociology, anthropology, political science, and economics have rejected the notion that violence is the sole result of biological factors and uncontrolled reflexes inherent to certain individuals or groups. Instead, there is broad theoretical consensus in acknowledging that economic, social, and cultural factors underlie violent behavior (Cavanaugh [2012]). The objective of this dissertation is to contribute to this scientific inter-disciplinary effort by empirically uncovering the economic and cultural determinants of several manifestations of violence. To do so, each chapter builds on theories of violence proposed by social scientists and tests their predictions using rigorous applied-microeconometric tools and newly-assembled datasets.

Chapter 1 opens the investigation of the determinants of violence by studying a widespread form: intimate-partner violence (IPV) against women. The main finding is that violence within the home can be traced back to colonial history. Using Cameroon as a laboratory, this chapter highlights that institutions that favor women's economic empowerment can have unintended

consequences. Consistently with sociological theories of male backlash, IPV emerges as a tool to reinstate a cultural norm of male dominance and female dependence. While Chapter 1 keeps the cultural setting constant, Chapter 2 shows that different cultural norms about gender can directly explain sexual violence in armed conflict, a brutal yet understudied form of gender-based violence. Some ethnic actors, when involved in a conflict, systematically engage in acts of rape, sexual slavery, or sexual torture. Other actors, instead, never do so or do so only sometimes. We improve the understanding of this puzzling phenomenon by showing that conflict-related sexual violence increases with the cultural distance in gender norms between the combatants. For the first time in the economics literature, this chapter demonstrates that cultural differences between ethnic groups can explain how violence unfolds once a given conflict starts. Building upon existing theories of conflict, Chapter 3 empirically shows that cultural differences between ethnic groups can also explain whether or not conflict occurs in the first place. Culturally distant ethnic groups within the same country hold different preferences over public policies. In turn, this triggers ethnic civil conflict over government power.

All three chapters of this dissertation are self-contained, they have their introductions and can be read independently. Each chapter has its own appendix and all appendices are included after Chapter 3. The bibliography containing all references can be found at the end of this dissertation. The remainder of this section provides a summary of each chapter.

Colonialism and Female Empowerment: A Two-Sided Legacy. Chapter 1, coauthored with Helmut Rainer, investigates the long-term effect of colonialism on the lives of women, including their exposure to IPV. While much is known on how colonial history has shaped development through institutions, legal systems, and infrastructure, the literature studying the link between colonialism and women's status remains scarce. This is surprising, given that, by their very nature, colonial institutions have profoundly impacted gender relations.

Cameroon is a particular case in point and provides a unique laboratory to shed light on this question. Between 1919 and 1961, Cameroon was divided between France and the United Kingdom. The two colonial regimes opened up divergent economic opportunities for women: unlike women in French territories, those under British colonial rule were economically empowered through access to paid jobs. The French and British territories were delimited by an arbitrary border, drawn up in Europe, that cut across politically, economically, culturally, and geographically homogeneous regions. In 1961, Cameroon was re-united into an independent state. We use the former Anglo-French border within today's Cameroon in a geographical regression discontinuity design. Using the Demographic and Health Survey (DHS), we compare women on the British side of the former colonial boundary to women on the French side. The first finding is that, after more than 40 years from Cameroon's independence, women on the former British territories are more likely to have access to paid employment, to own property, and to have a larger say in household purchases. In other words, women are more likely to be economically empowered.

This finding allows us to take a step forward and investigate a much-debated question within the gender economics literature. Does female economic empowerment affect women's exposure to IPV? From a theoretical standpoint, whether female economic empowerment decreases or increases IPV is unclear. On the one hand, household bargaining models predict that better economic opportunities for women reduce IPV by leveling the balance of power between partners. On the other hand, sociological theories of male backlash suggest that women's economic empowerment increases IPV: when gender roles and power relations are redefined, men resort to violence to reinstate a culturally prescribed norm of male dominance and female dependence.

We find that women in former British territories are not only more likely to be in paid employment, but also to be IPV victims. We present three additional sets of results and find evidence consistent with male backlash. First, we show that the British colonizer effect on IPV does not occur independently of the effect on female employment. Women who, due to British colonial rule, are more likely to be in paid employment are precisely the same women who face an increase in IPV victimization. Second, the entire colonizer effect on IPV is explained by an increase in women's joint probability of being victimized and with a partner who objects to their employment. Third, we rule out the IPV effects are driven by alternative channels like male employment, male occupations, male alcohol consumption, or education.

Taken together, these findings shed light on one of the ways IPV against women manifests. When women's economic status improves, men resort to violence to reinstate a culture of male authority and control over women. From a policy perspective, these findings highlight that enforceable laws offering women direct protection from IPV could reduce backlash as a byproduct of female economic empowerment.

Cultural Distance and Conflict-Related Sexual Violence. Chapter 2, coauthored with Ana Tur-Prats, examines the relationship between a specific component of culture—ethnic-based gender norms—and conflict-related sexual violence, an understudied technology of war and a form of violence predominantly targeted against women. This brutal form of violence includes, among others, acts of rape, sexual slavery, and forced prostitution. It is widespread in conflicts all over the world, but not ubiquitous: some conflict actors constantly perpetrate sexual violence, some do so only sometimes, and others never do so. In this chapter, we shed light on why this is the case.

We advance and test the hypothesis that sexual violence perpetrated during armed conflict is explained by cultural distance in gender norms between the combatants. We find that sexual violence is driven by a clash of conceptions on the appropriate role of men and women in society: sexual violence increases when the perpetrator is more gender-unequal than the victim, but not vice-versa. When confronting a more gender-equal opponent, perpetrators might feel threatened by the relatively better position of women in the opponent's society and thus resort to a gender-based form of violence to reinstate their own cultural ideal. Conversely, combatants that encounter a more gender-unequal society might not experience any menace or necessity to react against the different roles of men and women in the opponent's society.

The empirical analysis relies on a novel dyadic dataset containing information on the ethnic identity of armed actors and their use of sexual violence in all African ethnic conflicts fought between 1989 and 2009. To measure ethnic-specific gender norms, we exploit ethnographic information on the role of men and women in ancestral societies. We identify nine ethnic traits that, according to well-established interdisciplinary literature, relate to anthropological notions of gender equality or inequality, and group them into a single index capturing ethnicities' deep-rooted gender norms. Using survey data, we show that the index is a powerful predictor of an ethnic group's gender norms, as well as other measures of gender inequality. The large within-country variation in our index speaks in favor of the use of ethnic-specific gender inequality indices as opposed to the existing country-wide indices, especially in ethnically diverse regions of the world like Africa. Finally, to test whether sexual violence in conflict increases with the cultural distance in gender norms between the combatants, we exploit the dyadic structure of the data and run a specification in the spirit of a gravity equation.

This chapter uncovers a gender-based explanation for a gender-based form of violence. Cultural distance in gender norms does not explain general violence within a conflict and distances in other cultural dimensions unrelated to gender do not explain conflict-related sexual violence. We contribute to the literature on gender-based violence by studying sexual violence in armed conflict, which has been largely overlooked within the economics literature, and by constructing a novel gender inequality index at the ethnicity level. More broadly, this chapter contributes to the conflict literature by showing, for the first time, that ethnic differences can explain how violence within a conflict manifests. Finally, by focusing on cultural distances between belligerents, this chapter demonstrates that conflict is a function of the characteristics of all ethnic groups involved, and not only of the characteristics of one side, as highlighted in previous literature.

Cultural Distance and Ethnic Civil Conflict. Chapter 3 shows that cultural differences can explain not only the *intensive margin* of violence—i.e., how violence unfolds once conflict takes place, as in Chapter 2—but also the *extensive margin* of violence—i.e., whether ethnic groups engage in conflict or not in the first place. In this chapter, I study civil conflicts, the most widespread and deadly form of war since World War II. Roughly half of all civil conflicts worldwide have been fought along ethnic lines. Cross-country studies found that ethnically diverse countries tend to be more prone to conflict when compared to ethnically homogeneous ones. However, the reasons why some ethnic groups engage in conflict—that is, rebel against the central government—and others do not remain poorly understood.

I address this question by generating a novel dataset that allows me to move the empirical analysis of diversity and conflict from the country level to the ethnicity level. I examine how cultural distance—defined as differences in beliefs, values, and preferences—between each ethnic group and the ethnicities forming the central government affects a group's decision to rebel. I find that larger cultural distance to the government, measured using linguistic distance, increases conflict over government power, but decreases conflict over territory. Since every group in a country is subjected to public policies provided by the government, different preferences over

public policies give rise to disagreement and ultimately conflict. Instead, ethnic groups with different preferences and tastes are less likely to compete over the same type of territory or resource.

The link between cultural distance and conflict holds across three alternative research designs, each addressing specific identification challenges. The first strategy leverages within-ethnicity variation in cultural distance to the government resulting from power transitions between ethnic groups over time. The second strategy, a triple difference-in-differences design, focuses on groups that were partitioned across countries during the Scramble for Africa and are thus exposed to different governments in different countries. Third, to address endogeneity concerns, I use a novel instrumental-variables approach. As an instrument for cultural distance, I exploit the Bantu expansion, a prehistoric cultural shock that changed the culture of some parts of sub-Saharan Africa but not others.

To shed light on mechanisms, I use survey data and find that respondents are more likely to oppose a wide range of current government policies if they belong to ethnicities that are culturally distant from the ethnic groups in power at the time of the survey. Taken together, the analysis in this chapter speaks against the widespread idea that ethnic conflict is the sole result of insurmountable ethnic hatred between groups, as cultural distances only trigger a specific type of dispute, i.e., conflict over government power. This suggests that a mix of public policies accommodating diverging preferences might be powerful in alleviating conflict in multicultural societies.

Chapter 1

Colonialism and Female Empowerment: A Two-Sided Legacy*

1.1 Introduction

Colonial history is widely recognized as crucial for development paths, but what has been its role in shaping the lives of women? A significant body of research has established colonialism as an important determinant of long-term economic development, as it affected political institutions (Acemoglu *et al.* [2001]), legal systems (La Porta *et al.* [1997, 1998]), and public investments in areas such as education, health, and infrastructure (Huillery [2009]). Yet, there exists little well identified evidence about colonialism's long-lasting impact on women's economic, social, and health status.

We address this question by utilizing a natural experiment of history in Cameroon. At the end of World War I, after three decades as a German colony, the western territories of today's Cameroon were divided between France and the United Kingdom under a League of Nations mandate. The French and British territories were delimited by an arbitrary border, drawn up in Europe, that cut across politically, economically, demographically and geographically homogeneous regions. Historical records provide evidence that French and British colonial practices affected women on the two sides of the border in significantly different ways. Several British colonial policies created new educational and occupational opportunities for women. For example, girls benefited from a protestant education system that aimed at educating both girls and boys, and women gained opportunities to earn cash wages in the export oriented agriculture sector, under the same conditions as their male counterparts. In contrast, the French colonial practice in these domains centered around educating a small administrative elite and investing

*This chapter is based on joint work with Helmut Rainer. A version of this Chapter has been published in the *Journal of Development Economics*, Volume 151, Article 102666 in June 2021, doi: `https://doi.org/10.1016/j.jdeveco.2021.102666`.

in the male employment dominated infrastructure sector, to build railways and roads. In 1961, the political evolution of Cameroon culminated in the reunification of its divided colonies into an independent state.

This setting provides us with a unique opportunity to examine whether and how policies promoting female empowerment affect women's lives in the long-term. We focus on two questions: Did the divergent opportunities that opened up for women when Cameroon was split into a French and a British colony have a long-term impact on female participation in activities outside the household, especially in the labor market, that persists until today? Given that women's employment status has been linked to domestic violence in numerous settings worldwide, does Cameroon's colonial past continue to influence women's treatment within the household nowadays, especially in terms of intimate partner violence (henceforth, IPV)?

To get at these questions, we use the historical Anglo-French border in present day's Cameroon in a spatial regression discontinuity (RD) design. Our analysis draws upon two repeated cross sections (2004, 2011) of the Cameroon Demographic and Health Survey (DHS), which contains geolocated household survey data and a domestic violence module, completed by one randomly selected woman per surveyed household. Because the validity of our spatial RD design rests on continuity of all factors besides treatment at the historical Anglo-French border, we match ethnic groups in the DHS with information on ancestral anthropological and cultural group characteristics from Murdock's Ethnographic Atlas. Specification checks show that these ancestral characteristics, some of which have been shown to correlate with women's current outcomes (Alesina *et al.* [2021]), do not vary across the historical Anglo-French border. We also provide evidence suggesting that selective migration across the treatment threshold poses no threat to identification.

We present five sets of results. First, we show that the British colonial rule had a two-sided legacy for women that is still visible today. On the one hand, there remains a persistent impact of the historical Anglo-French border on female participation in the labor market. In particular, women in former British territories have a 24 percentage points higher propensity to be in paid employment than their counterparts in former French territories. In terms of magnitude, this is a large effect considering that the mean female paid employment rate throughout the examined region is 62%. Consistent with the British colonizer effect on female employment, we find that women on the British side of the border are also more empowered in terms of control over household resources and property ownership. On the other hand, we estimate that women on the British side of the historical border face a 10 percentage points higher risk of (past year) spousal violence than their counterparts on the French side. Just as the border effect on female paid employment, this effect is large in magnitude as it compares with a mean prevalence of IPV of 28% throughout the region examined. We demonstrate that the magnitude and precision of these estimates are robust to several alternative specifications.

Second, we ask if the two colonizer effects on female employment and IPV are explained by the same or two different subsamples of women. The latter is a possibility as the partition of Cameroon and its later reunification induced a multiplicity of treatments: besides its impact on

women's economic opportunities, the identity of the colonizer has likely affected many dimensions of life in Cameroon, some of which might be relevant for domestic violence in isolation.[1] We show that the two separate border effects on female employment and IPV are not explained by two separate subsamples of women, with one subsample driving the employment effect and the other driving the IPV effect, but by the same subsample of women. In order words, those women who, due to the legacy of British colonial rule, are more likely to be employed are also those who experience an increased likelihood of being victims of IPV.

Third, we conduct supplementary analyses to explore some possible explanations for our findings. The result that the British colonial rule caused a persistent increase in both female employment and IPV, and that these two effects are driven by the same subsample of women, is inconsistent with household bargaining models. These models predict that increased economic opportunities for women reduce IPV by leveling the balance of power between partners (Aizer [2010]). By contrast, our results are consistent with theories of male backlash, whereby men resort to violence when their partners' economic opportunities increase, in order to reinstate a culture of male authority and control over women. Supporting this interpretation, we show that women of the British side of the historical border have a 12 percentage points higher propensity of having a partner who objects to their employment, and that this effect is entirely explained by the subsample of women driving the British colonizer effect on female employment and IPV.

Fourth, while the evidence we provide is suggestive of male backlash (i.e., a causal chain from British colonial rule to higher female employment to increased IPV risk), there might be colonizer effects on third factors driving female employment, domestic violence, or both. Thus, we explore observed individual and household level factors that the literature highlights as such potentially important drivers: male employment, male occupation, male alcohol consumption, education, number of children, age, practice of polygamy, and exposure to conflict. Our RD results show that these factors vary smoothly at the historical Anglo-French border.

Fifth, a potential concern with the interpretation of our results arises if self-reporting of IPV is affected by women's empowerment and, in our setting, by different colonial practices in the past. It could be argued that empowering women changes their attitudes to gender equality and, hence, increases the likelihood of reporting IPV in their daily lives. Agüero and Frisancho [2017] use an experimental approach based on indirect questioning techniques to assess the extent of truthful IPV reporting in the DHS. They find no evidence of IPV misreporting for low-educated women, but highly educated women appear to underreport IPV victimization. This suggests that, if anything, empowered women—i.e., those who do not fit the typical victim stereotype—face larger disutility of being exposed as victims and therefore underreport IPV. Building on this finding, if IPV reporting of empowered women is a threat to our interpretation of the results, we would expect a positive individual level correlation between IPV and different measures of female empowerment other than employment. Multivariate regressions based on our estimation sample show this not to be the case: there is a negative correlation between

[1] For example, the colonizers may have brought with them different cultural norms about the acceptability of violence against women or institutional features conducive for women's bargaining power.

domestic violence and a woman's education, which is inconsistent with the notion that IPV reporting increases with a woman's empowerment.

Taken together, the British colonizer effect on female employment and IPV, the individual-level links between own employment, IPV and partner controlling behavior, and the null results for third factors plausibly affecting female employment and IPV shed some light on competing theories of domestic violence. The results are incompatible with economic theories of household bargaining that explain better outcomes for women inside the household with better female opportunities in the outside labor market; they are, however, consistent with theories of male backlash, i.e., by the idea that men might resort to violence when their partners' economic status improves in order to reinstate a culture of male authority and control over women. From a policy perspective, one might read our results as a cautionary tale: in low-income country settings, it would be too quick to equate increased female opportunities in the labor market with universally better outcomes for women, especially inside households. For this to materialize, enforceable laws that offer women direct legal protection from IPV and/or the opportunity to divorce from abusive partners would seem a (so far often not existing) precondition.

To our knowledge, we are among the first to exploit Africa's colonial past to study its long-term impact on women's empowerment outcomes. Our work builds upon and connects three distinct areas of inquiry: a growing literature in economics on the causes of IPV, work on female empowerment and changes in women's outcomes, and a large body of research on the legacy of colonialism. Let us first synthesize the IPV literature, with a focus on studies that examine the impact of women's labor market participation and opportunities in developing countries. Based on individual-level data from 30 sub-Saharan countries, Cools and Kotsadam [2017] find a strong positive association between female employment and IPV, especially in areas with a high level of acceptance of wife beating. Bhalotra *et al.* [2018] use business cycle variations across 31 developing countries to examine how IPV responds to improved employment prospects for women. A key result of this study is that an increase in female unemployment rates is associated with a decrease in women's probability of IPV victimization. The findings of both Cools and Kotsadam [2017] and Bhalotra *et al.* [2018] are consistent with male backlash.

A number of papers have studied the extent to which female empowerment translates into changes in women's outcomes. A study that is closely related to ours is that by Anderson [2018], who provides evidence based on the DHS supporting the following argument: In sub-Saharan Africa, British common law is associated with weaker female marital property rights than French civil law. This, in turn decreases women's bargaining power and ability to negotiate safe sex practices, and results in higher HIV vulnerability among women in common law countries. Although the bargaining power argument highlighted by Anderson [2018] could, in principle, also explain our findings, it does not have bite in the case of former British and French Cameroon. Among sub-Saharan civil law countries practicing community marital property, Cameroon is the only case where married women are *not* entitled to some of the marital property in the case of divorce (Hallward-Driemeier *et al.* [2012]). Thus, although elements of common law are still applied in former British territories of Cameroon, women on the two sides of the

historical border face the same legal outside options to marriage. In our data, we find indeed no evidence of women in former French Cameroon having more bargaining power than their British counterparts (e.g., in terms of their say in intra-household decision-making). A small set of papers provides well-identified evidence that increased opportunities for women in the labor market improve women's outcomes in the long-term: it has decreased the number of "missing girls" in rural China (Qian [2008]) and improved girls' access to education in India (Munshi and Rosenzweig [2006]). An excellent survey of this and related studies is contained in Duflo [2012].

There is now a body of research, too large to review here but synthesized by Nunn [2014b], examining how critical events in history, and subsequent institutional paths, matter for current economic development. Our paper relates to a subset of this literature that explores the long-term impact of Europe's colonization of Africa, including the role of investments in education and religious missions (Bertocchi and Canova [2002], Huillery [2009], Nunn [2010], Cogneau and Moradi [2014]). Two papers in the literature exploit, as we do, the Anglo-French division of Cameroon after WWI. Lee and Schultz [2012] study its impact on household wealth, while Dupraz [2019] focuses on education. Although these authors exploit the same setting as we do, the questions they address differ markedly from ours. Most importantly, our paper exploits that the colonial division of Cameroon involved a stark gender divide, with French and British colonial practices affecting women's education and employment opportunities in significantly different ways. As such, we believe that the historical partition of Cameroon offers a unique setting to study the long-term impact of female economic empowerment on intra-household outcomes. An important methodological antecedent to our setup is Dell's [2010] study of Peru's mining *mita*, which set the standard for the use of border discontinuity designs in historical settings.

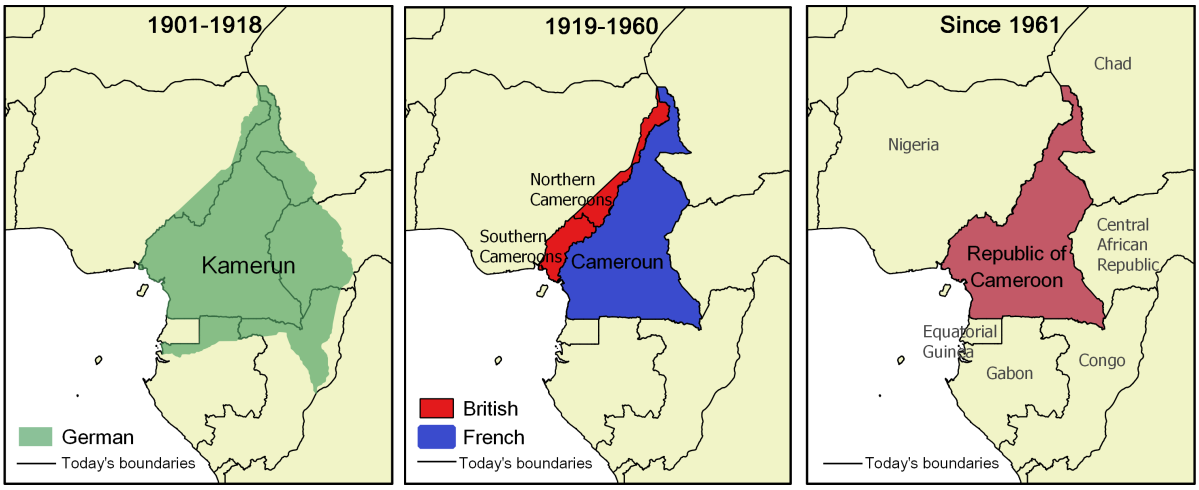
In the next section, we provide background on the history of Cameroon. Section 3 describes the data sources we use. Section 4 explains our estimation strategy and provides specification checks. Section 5 tests the long-term effect of British versus French colonization on women's outcomes and probes the robustness of the findings. Section 6 examines mechanisms. Section 7 provides a discussion on self-reporting of IPV. Section 8 concludes.

1.2 Historical Background

1.2.1 The Colonial Partition

The boundaries of today's Cameroon are the result of a sequence of historical events that originated at the end of the 19^{th} century, when Germany established the protectorate of Kamerun. In addition to the lands of today's Cameroon, German dominations included also portions of Nigeria, Chad, Central African Republic, Congo and Gabon (Figure 1.1, left panel). The German presence in Cameroon saw the construction of roads and railway lines, and the introduction of a system of forced labor to recruit the necessary manpower for these projects (Le Vine [1964], p. 104).

Figure 1.1: *Evolution of Boundaries in Cameroon*



Sources: Historical boundaries: Gifford and Smith [1967] and Louis and Gifford [1971]. Contemporary boundaries: http://www.maplibrary.org/library/stacks/Africa/index.htm.

With the outbreak of World War I, the Allies invaded German territories in Africa. Cameroon was first provisionally partitioned into two parts by the so-called Picot line, which was decided upon when the British representative Lancelot Oliphant asked the French representative George Picot to draw a line on the map of Kamerun during a meeting in London in February 1916 (Yearwood [1993], p. 225). This provisional arrangement was then ratified at the end of the war in 1919, during the Paris Peace Conference. As part of the Treaty of Versailles, the League of Nations assigned to Britain and France a mandate to govern the newly divided territories. The provisional Picot line—with only minor changes—became the international boundary between the French dominated Cameroun in the East and the British Cameroons in the West, the latter divided into Southern and Northern Cameroons respectively (Figure 1.1, middle panel). In addition to the fortuitous historical circumstance that saw the territories of former Kamerun partitioned, it was also apparent to many officials at the time that the boundary that cut through Cameroon was arbitrary in nature; one particularly poignant comment read:

> *"The boundaries of the zones [...] are haphazard and, as a permanent arrangement, would be quite intolerable. They cut across tribal and administrative divisions, take no account of economic conditions, and are in any way objectionable."*—Lord Alfred Milner, British Colonial Secretary, 1918-1919 (cited in Louis [1967] and Lee and Schultz [2012])

Figure A.1 further supports this view: the colonial border cut through both agro-ecological zones and pre-colonial ethnic settlements. A post-independence glance at the partition of Cameroon well summarizes the subsequent four decades of history:

> *"Arbitrarily sundered into three parts, the territory lost whatever unity it had achieved during the [German] protectorate. The two Cameroons under separate administrations moved off in different directions, propelled by the force of colonial policy often*

> *diametrically opposed to one another. The artificial bisection of the territory created the reality of two distinctly different Cameroons, with different social, economic, and political traditions."*— (Le Vine [1964], p. 35)

Cameroun gained independence from France in January 1960, and became the Cameroun Republic. In 1961, a plebiscite took place in North and South British Cameroons in order to establish whether the areas would join Nigeria or the newly-created Cameroonian Republic. While voters in Northern Cameroons opted for being annexed to Nigeria, Southern Cameroons joined the French part with a majority of around 70% of votes (Le Vine [1964] p. 212). In 1972, the federal system characterizing the new republic was abolished in favor of the United Republic of Cameroon.

1.2.2 Colonial Policies

In the almost 42 years of colonial partition, there are various dimensions along which the British and the French administration differed. We here provide a summary of what characterized the institutional and legal setting, the education system, and the labor market policies, focusing on how Anglo-French differences in these spheres affected the position of women in society. Section A.1 in the Appendix provides additional details and references to historical records.

From the institutional viewpoint, the British administration was characterized by the practice of indirect colonial rule, which incorporated native chiefs into the local political system. Female participation in political life was formalized by the Southern Cameroons Electoral Regulations in 1957, which stated that "[women] may vote and stand as candidates for election under the same conditions as men" (Government of the United Kingdom [1958b], par. 635 and 991). Women were regarded as men in front of the law and were entitled to acquire and hold their own property, as prescribed in the Married Women's Property Acts introduced in the United Kingdom in 1882. As it was the case in their other colonies, the British implemented a legal system based on common law, which is still applied in the Anglophone regions of today's Cameroon.

On the other side of the border, the French applied a system of direct rule based on the policy of assimilation, guided by the mission to civilize Africans according to Western principles and institutions (Le Vine [1964], p. 91). The French implemented a civil law system, combined with a practice of legal differentiation, which applied different policies and standards to Africans according to their advancement in the "evolution" towards the French ideal. Until 1946, separate legal systems were assigned to *citoyens*, assimilated to European law, and to the so-called *sujets*, subject to native customary rules. The evolution to the privileged *citoyens* status was precluded to women, who faced restricted opportunities for formal employment and education.

Recent work by Anderson [2018] has highlighted how the British legal system based on common law, by establishing separate marital property, generated weaker female marital property rights and low bargaining power for women, compared to the French civil law system. Civil law countries in sub-Saharan Africa prescribe a community marital property regime, which provides equal protection to women upon divorce by evenly splitting property between spouses and rec-

ognizing women's non-monetary contributions. However, this does not apply to former French Cameroon, as documented in the Women's Legal and Economic Empowerment Database for Africa (Hallward-Driemeier *et al.* [2012]). Among civil law countries practicing community marital property, Cameroon is the only case where women in statutory marriages are *not* entitled to some of the marital property in the case of divorce.[2] The wife's non-monetary contributions are recognized, but marital property is *not* divided equally upon divorce, exactly as in the vast majority of common law countries with separate marital property regimes. Thus, the distinction between common and civil law countries in sub-Saharan Africa, as exploited by Anderson [2018], does not apply to the case of former British and French Cameroon. This is confirmed by the fact that women in former British territories, where elements of common law are still applied today, detain higher bargaining power (in terms of their say in intra-household decision-making) than their counterparts in French territories (see Table A.4).

Turning to education policies, the two colonial powers markedly differed in the system they promoted. The British allowed English speaking Protestant missions to monopolize the supply of education. Across sub-Saharan Africa, Nunn [2014a] has shown that Protestant missionary activities had a long-lasting impact on the education of girls, as opposed to Catholic schools, which exerted a greater impact on male education in the long run. Protestantism encouraged girls' education, based on Luther's idea that women had to be able to read the Bible in order to go to heaven. The French, parallel to predominantly Catholic mission schools, instituted a public system aimed merely at the education of a restricted elite of native young men, who would become functional to the colonial administration.

Finally, the two colonial administrations implemented divergent labor policies. After the partition of Cameroon, the British abolished the German practice of forced labor, and introduced cash wages. Plantation labor opened up unprecedented employment opportunities for the local population, and in particular for women. The increased participation of women in the labor market is documented in the British colonial report of 1958, which highlights the newly gained opportunities for women to earn cash wages under the same conditions as their male counterparts (Government of the United Kingdom [1958b], par. 637). In order to efficiently manage plantation labor, the British administration established the Cameroon Development Corporation in 1947. During colonial times, the Corporation employed more than half of wage earners in Southern Cameroons (Government of the United Kingdom [1958b], par. 650). It continued to exist also after independence, specializing in the production of tea, palm products, bananas, cocoa and rubber for export purposes. Nowadays, it constitutes the second largest employer after the state of Cameroon.

On the other hand, the French administration kept the extractive labor system introduced by the Germans. Under the form of a *prestation*, it consisted in the obligation to supply ten days of free labor a year for Cameroonian men of *sujet* status. The main goal was to recruit the necessary workforce to implement large scale infrastructure projects.

[2]The same applies to women in customary marriages and consensual unions.

There are no official statistics on female employment during colonial times. However, post-independence data show how a large gap in female employment in former British and French dominated zones existed when first official statistics became available in the 1970s. Figure 1.2 displays male and female employment rates in the former British and French territories respectively, using data from the 1976 Cameroonian census.[3] As can be seen in Panel (a), male employment rates did not differ across the two regions. However, when turning to female employment, the picture changes significantly: women in former British areas were 7 percentage points (or 18%) more likely to be employed, compared to women in the neighboring territories subjected to French domination. This is in line with the just-described historical evidence that women under the British colonial rule faced better employment opportunities than those under French rule.

Panels (b) and (c) of Figure 1.2 further split the sample into two age cohorts: (i) individuals aged 31-49, i.e. those who were 15 or older before independence and thus entered the labor market during colonial times; (ii) individuals aged 15-30, a younger generation that entered the labor force after Cameroon's reunification. An interesting pattern emerges: while no large differences in male employment arise when observing the two age groups separately, the employment gap favoring women in former British areas amplifies (i.e. almost doubles) when moving from one age cohort (31-49) to the next (15-30). This already suggests that the divergent colonial practices could have produced persistent labor market effects for women that are passed down through generations. We will verify this conjecture by showing that the cross-border difference in female employment still persists today, in almost the same magnitude as in the 1970s.

1.2.3 Contemporary Domestic Violence Legislation

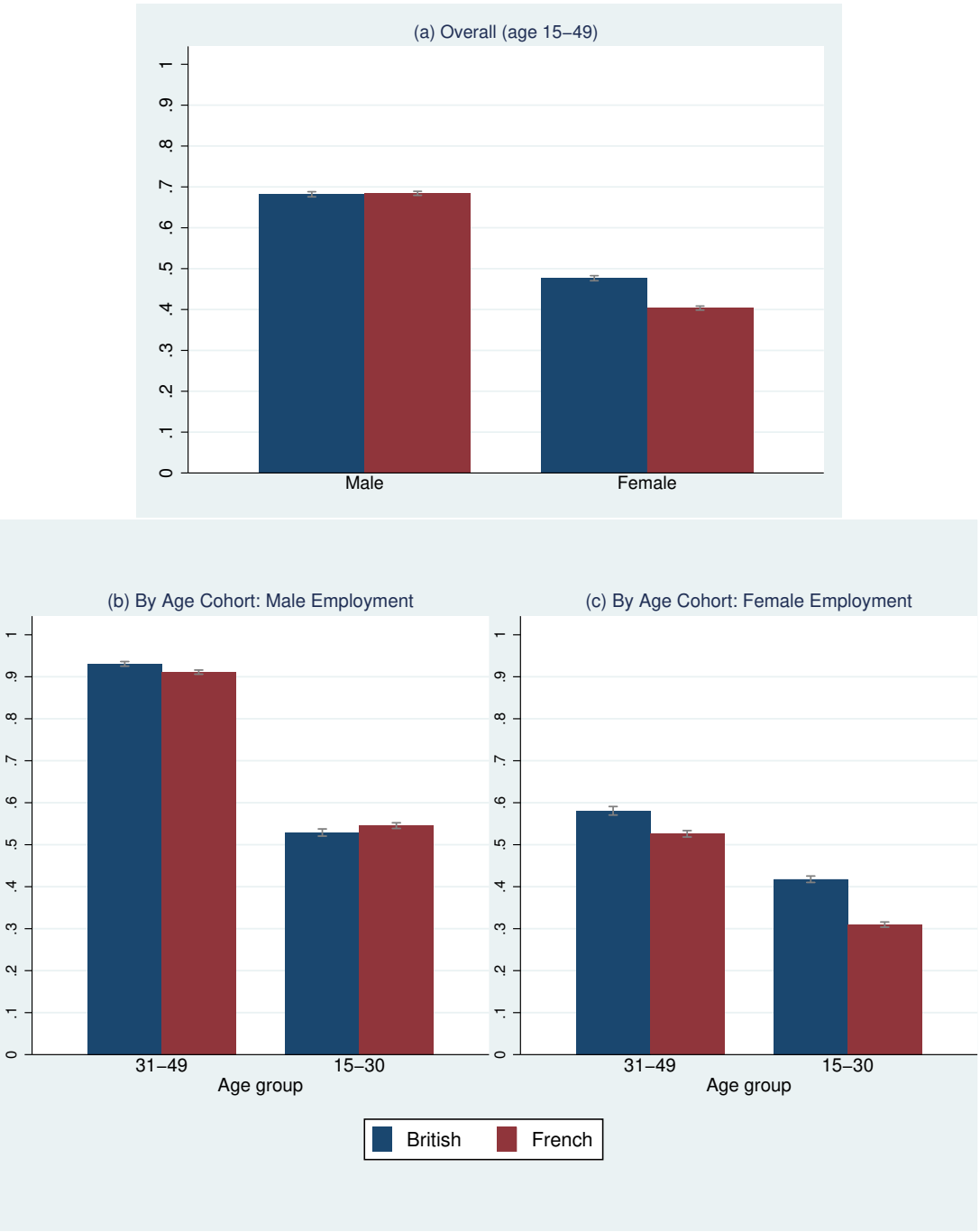
One of the outcomes considered in this study is IPV. In contemporary Cameroon, domestic violence is not recognized as a crime (SIGI [2019]). Since there is no legislation against domestic violence, victims willing to report instances of abuse have to rely on the general law of assault. The Cameroonian Penal Code prohibits instances of assault, including rape. In this latter case, perpetrators are formally allowed to marry the victim and avoid persecution. Under customary law, spousal rape is generally not considered an offense, as it is assumed that a married woman consents to sexual intercourse with her husband (United States Department of State [2017]). Finally, exposure to spousal abuse is not a legal ground for divorce in Cameroon (SIGI [2019]).

1.3 Data

Our empirical analysis of IPV in Cameroon builds upon two recent waves of the Cameroon Demographic and Health Survey (CDHS 2004 and CDHS 2011). These two waves not only contain rich socio-economic individual- and household-level information, but also a domestic

[3]For consistency with the rest of the analysis, we restrict the sample to individuals aged 15-49 residing in an *arrondissement* (the lowest administrative unit in Cameroon) whose capital lies within 40 km of the Anglo-French border.

Figure 1.2: *Female and Male Employment in the 1976 Census*



Notes: The graphs compare female and male employment rates in former British and French areas, using data from the 1976 Cameroonian census (the IPUMS-I 10% extract: https://international.ipums.org/international/). The sample includes individuals between 15 and 49 years of age, resident in an *arrondissement* (the lowest administrative unit in Cameroon) whose capital lies within 40 km of the Anglo-French border. Panel (a) reports overall employment rates by gender and by region of residence. The sample includes 63,090 women (39.8% on the British side) and 54,876 men (38.9% on the British side). Panels (b) and (c) report female and male employment rates separately for two age groups: between 31-49 years, i.e. comprising individuals that were 15 or older (and therefore entered the labor market) before independence; 15-30, i.e. comprising individuals that entered the labor market after independence. Sample sizes are as follows. Women 31-49: 25,539 (35.5% British); women 15-30: 37,551 (42.7% British); men 31-49: 20,860 (39.0% British); men 15-30: 34,016 (38.9 % British). The vertical light-gray lines represent 95% confidence intervals.

violence module which was administered to one randomly selected woman per surveyed household. Measuring a sensitive issue like IPV poses important challenges. Administrative records (i.e. from the police, health centers or social services) severely underestimate the phenomenon, which has been explained by the exposure costs imposed by reporting spousal abuse to the authorities (Agüero and Frisancho [2017]). Palermo *et al.* [2013] show that, in 12 sub-Saharan African countries where the DHS survey was implemented, only 6.2% of women who experienced violence reported it to formal services. This evidence points to the need to rely on self-reported measures of IPV such as those provided by the DHS.

The advantage of DHS data is at least twofold. First, the domestic violence module meets the requirements set out by the World Health Organization to (i) guarantee confidentiality and (ii) attenuate the phenomenon of underreporting (World Health Organization [2001]). Rigorous privacy and ethical protocols are applied during the interview: fieldworkers are required to guarantee a safe environment, making sure that women are alone when asked questions about IPV. Second, and equally importantly, its design restricts the room for discretion of the respondent to a minimum by avoiding generic and subjective questions such as "Have you ever experienced domestic violence?" and rather asking about specific acts of violence. These questions include: *Did your husband/partner: (i) push you, shake you, or throw something at you? (ii) slap you or twist your arm? (iii) punch you with his fist or with something that could hurt you? (iv) kick you or drag you? (v) try to strangle you or burn you? (vi) physically force you into unwanted sexual intercourse?*. This procedure reflects a revised version of the Conflict Tactics Scales as first advocated by sociologists (Straus [1979], Straus *et al.* [1996]), and is considered by many social scientists as the gold standard for survey data collection on IPV.

We construct two measures of IPV. The first is a dummy variable indicating whether the respondent experienced any of the just listed forms of IPV by her partner during the 12 months leading up to the interview. The second is an index ranging between 0 and 6 and counting the different types of physical and/or sexual aggression to which the respondent was exposed in the past year. To investigate potential channels, we exploit a rich set of individual- and household-level information, including female and male employment, male occupations, a household wealth index, years of education, fertility, women's say in household purchasing decisions, ownership of assets, male alcohol consumption and polygamy.

For our estimation strategy, which rests on a geographic regression discontinuity design, it is of paramount importance to geo-locate all surveyed households. In the DHS, the location of households from the same enumeration area (henceforth, cluster)—typically villages in rural areas or neighborhoods in cities—are aggregated to a single point coordinate. To ensure respondent confidentiality, this coordinate is then randomly displaced through use of the Global Positioning System (GPS) coordinate displacement process: urban clusters are displaced a distance up to two kilometers (0-2 km) and rural clusters are displaced a distance up to five

kilometers (0-5 km), with a further, randomly-selected 1% of rural clusters displaced a distance up to 10 kilometers (0-10 km).[4]

For identification checks—in particular, to probe the the randomness of the historical Anglo-French border—we exploit the Ethnographic Atlas (EA) coded by George Peter Murdock [1967] and updated by Nunn and Wantchekon [2011]. The EA provides rich ethnographic information for 1,265 societies around the world as of the end of the 19th century. For Africa, the EA sheds light on socio-economic conditions, settlement patterns and family arrangements prior to European colonization. In our setting, this information provides a unique opportunity to examine the extent to which ethnic groups inhabiting areas located very close to the colonial boundary are similar in terms of ancestral characteristics (e.g., use of the plough, endogamy, marriage payments, polygamy) that have been shown to correlate with contemporary levels of violence (Alesina *et al.* [2013]). We also check for geographic similarities on the two sides of the historical border by exploiting information on agro-ecological zones as classified by the Food and Agriculture Organization (FAO) and collected in a finely spaced grid by the International Food Policy Research Institute (IFPRI) [2015]. Last, we exploit Nunn and Wantchekon's [2011] ethnicity-level data on trans-Atlantic slave shipments in the 19th century, which have been shown to be a historical determinant of long-run development, trust, and female participation in the labor market (Nunn [2008], Nunn and Wantchekon [2011], Teso [2018]).

Finally, we will investigate the role of conflict in explaining differences in IPV across the historical border by using the UCDP Georeferenced Event Dataset (UCDP GED), which collects events related to organized violence (state-based conflict, non-state conflict and one-sided violence) and locates them in both time and space, between 1989 and 2016 (Sundberg and Melander [2013]).[5]

1.4 Estimation Strategy and Specification Checks

1.4.1 Estimation

We exploit the former Anglo-French border as a discontinuity to identify the reduced-form effect of British versus French colonial rule on contemporaneous levels of IPV. Specifically, we compare outcomes for Cameroonian women located close to the British side of the boundary to their counterparts on the French side in a spatial regression discontinuity framework. Our basic regression follows Dell [2010] and Dell *et al.* [2018] and takes the form:

$$\text{DV}_{icbt} = \alpha + \gamma \text{British}_c + f(\text{geographic location}_c) + X'_{ic}\beta + \Phi_b + \tau_t + \epsilon_{icbt} \tag{1.1}$$

[4]Since this displacement procedure is random, it induces classical measurement error which would bias our estimates towards zero. Moreover, since our treatment variable—a dummy variable indicating whether a cluster was formerly subject to the British or French colonial regime—is constructed based on the cluster's province and not its coordinate, we do not run into the problem of assigning the "'wrong" colonial regime to it. We will be more specific about this below.

[5]Dataset available at: http://ucdp.uu.se/downloads/.

where DV_{icbt} is the outcome variable of interest for woman i in cluster (village) c along segment b of the boundary in survey year t. $British_c$ is an indicator equal to one if the observation i belongs to a village which was subject to the British rule until 1961 and equal to zero if subject to French colonization;[6] $f(\text{geographic location}_c)$ is the RD polynomial, which controls for smooth functions of geographic location of village c; it can be specified in two ways: either multi-dimensionally, with latitude and longitude; or with distance to the boundary, a more parsimonious mono-dimensional measure. We present results for distance to the boundary, latitude and longitude, or both as running variables. Following a recent contribution by Gelman and Imbens [2019], we use a local linear polynomial for our baseline specification, and show robustness for polynomials of higher orders (quadratic and cubic). X'_{ic} is a vector of geographic covariates containing distance to the capital and distance to Douala, the two largest urban areas. We control for Φ_b, a vector of border segment fixed effects, obtained by splitting the boundary in four segments of equal length. This rules out the possibility that treated units located at the extreme north are compared to counterparts at the extreme south. Finally, τ_t includes survey year fixed effects. Standard errors are clustered at the DHS survey cluster level.

We run Equation (1.1) for a restricted sample of individuals located close to the historical Anglo-French colonial boundary. As pointed out in Dell *et al.* [2018], the literature has not yet proposed an optimal bandwidth algorithm for regression discontinuity designs in setting like ours, where the running variable is multidimensional. Nevertheless, we calculate the data-driven optimal bandwidth as proposed by Calonico *et al.* [2019] using distance to the boundary as running variable and including the above-described covariates. We choose a window of 40 km around the boundary for our preferred specification. Figure 1.3 provides an illustration of DHS clusters within 40km of the colonial border. Reassuringly, our main results are robust to numerous alternative bandwidth choices.

1.4.2 Specification Checks

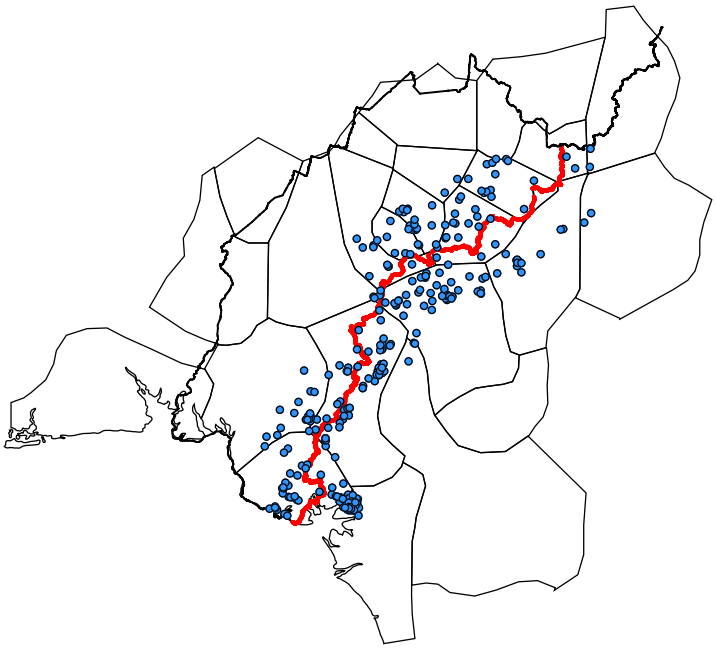
The coefficient γ can be given a causal interpretation if two identifying assumptions are satisfied. The first assumption (i.e., continuity) requires that all factors besides exposure to different colonial regimes vary smoothly at the historical border. In other words, our estimates would be biased if there were pre-existing cross-border differences in dimensions relevant to our outcome before Cameroon was split into a British and a French part. More specifically, a major concern would be if areas on the two sides of the border differed systematically in geographical features and ethnic characteristics that influence the incidence of IPV in the long-term.

Clearly, the continuity assumption is not entirely testable, but we are able to assess balance in baseline characteristics along two arguably important dimensions: geography and ethnicity.[7] Regarding geography, we first of all test whether areas around the colonial boundary are

[6] As mentioned above, to construct this indicator, we refrain from using latitude and longitude, as they are characterized by measurement error due to random displacement. Instead, we construct based on a variable indicating the province. This ensures that no cluster is erroneously located on either side of the border.

[7] Nunn [2014b] well summarizes the literature highlighting the role of geography in shaping long-term outcomes, especially economic development.

Figure 1.3: *Study Region, DHS Clusters and Colonial Boundary*



comparable in terms of climate and altitude. For a measure of climate, we match the DHS clusters to the FAO agro-ecological zones classification shown in Figure A.1, and then construct a binary variable indicating whether a cluster is located in a grid cell classified as cool/humid. Our elevation measure, which is included in the DHS, captures each cluster's altitude in meters. Columns (1) and (2) in Table 1.1 report coefficients obtained by running Equation 1.1 with climate and altitude—all measured at the DHS cluster level—as outcome variables. The small magnitude and insignificance of the estimates suggest no discontinuity of geographic features at the colonial border.

A large body of literature has investigated the role of ancestral ethnic features in shaping contemporary outcomes, in particular related to development (Michalopoulos *et al.* [2018], Michalopoulos and Papaioannou [2013]). A recent contribution to this literature has brought to light how some of these dimensions matter also for contemporary domestic violence. In particular, Alesina *et al.* [2021] show that part of the variation in IPV across sub-Saharan Africa can be explained by differential pre-colonial ethnic practices such as dependence on female- versus male-dominated economic activities (gathering versus fishing), marriage practices (endogamy versus exogamy, brideprice versus dowry, polygamy), and the use of the plough. If ethnic groups in areas close to the boundary already differed across these dimensions before the arrival of Europeans, the coefficient γ would be biased: it would capture the effect of pre-colonial ethnic characteristics rather than the impact of colonialism. As part of our test for the continuity assumption, we investigate balance of the just-described covariates in regions close to the colonial border using Murdock's Ethnographic Atlas. Figure A.2 shows that, throughout the region examined (and therefore across the colonial border), ethnic groups share the same ancestral characteristics in domains such as dependence on gathering, the use of the plough,

Table 1.1: *Balance Checks*

	Geography		Ethnic Characteristics (based on clusters' location)		
	Cool/Humid	Elevation	Dependence on Fishing	Practice of Endogamy	ln(1 + Slave Exports/Area)
	(1)	(2)	(3)	(4)	(5)
British ($\hat{\gamma}$)	-0.029	64.007	0.010	-0.079	0.203
	(0.089)	(88.677)	(0.025)	(0.083)	(0.148)
Mean Dep. Var.	0.287	705.9	0.127	0.967	0.266
Observations	2,030	2,030	2,030	1,936	2,030
Clusters	358	358	17	14	17

	Ethnic Characteristics (based on self-reported ethnicity)					
	Polygamy	Brideprice	Practice of Endogamy	Dependence on Fishing	Dependence on Gathering	ln(1 + Slave Exports/Area)
	(6)	(7)	(8)	(9)	(10)	(11)
British ($\hat{\gamma}$)	-0.071	-0.136	0.002	0.017*	0.553	-0.219
	(0.065)	(0.088)	(0.005)	(0.009)	(0.516)	(0.195)
Mean Dep. Var.	0.972	0.954	0.984	0.0588	2.944	0.456
Observations	1,928	1,928	1,402	1,879	1,879	1,928
Clusters	31	31	27	30	30	31

Notes: The unit of analysis is the survey respondent. The sample includes individuals in villages located within 40 kilometers from the border. *Cool/Humid* denotes whether the respondent belongs to a cluster located in an AEZ classified as tropic - cool/humid in the International Food Policy Research Institute (IFPRI)'s [2015] classification. *Elevation* is provided by DHS and measured at a cluster level. Ethnic characteristics are matched to respondents using their cluster of residence's location in columns 3-5. In columns 6-11, respondents' self-reported ethnicity is matched to one or more ethnic groups in the Murdock Ethnographic Atlas using the Linking Ethnic Data for Africa (LEDA) R package (Müller-Crepon *et al.* [Forthcoming]). When a respondent's self-reported ethnicity is assigned to multiple Murdock groups, the dependent variable is the average ethnic characteristic across all matched groups. *Polygamy*, *Brideprice*, and *Practice of Endogamy* are binary variables. *Dependence on Fishing* and *Gathering* are continuous variables that take values between 0 and 1, reflecting the extent to which a respondent's ancestral society relied on fishing or gathering for subsistence. All these variables are taken from Murdock [1967]. *ln(1+Slave Exports/Area)* is an ethnicity-level measure of the impact of the Atlantic slave trade in the 19th century, taken from Nunn and Wantchekon [2011]. Boundary segment fixed effects, distance to Douala and distance to the capital are present in all regressions. The estimated regressions use a linear polynomial in distance to the boundary as RD polynomial. Standard errors reported in parenthesis are at the DHS survey cluster level (columns 1-2) and at the ethnic-group level (columns 3-11). *** (**) (*) indicates significance at the 1% (5%) (10%) level.

the prevalence of brideprice, and the practice of polygamy. However, in terms of ethnic groups' reliance on fishing and the practice of endogamy, there is some variation across the region we examine. If this variation was systematically related to the colonial border, this would pose a threat to identification. In order to check this possibility, we first match individuals with the Ethnographic Atlas via their clusters' latitude and longitude. Columns (3) and (4) in Table 1.1 show that ethnic groups' dependence on fishing and practice of endogamy varies smoothly at the colonial boundary, i.e., that ethnic groups on the two sides of the border had balanced pre-colonial characteristics in these two domains.

The historical exposure to the slave trade is an important pre-colonial predictor of long-run development (Nunn [2008]) and contemporary trust levels (Nunn and Wantchekon [2011]). Moreover, Teso [2018] highlights how African women belonging to ethnic groups which were more affected by the trans-Atlantic slave trade are more likely to participate in the labor market today. The demographic shock—which generated a shortage of men and consequently a skewed sex ratio favoring women—altered the division of labor in society and shaped more equal gender-roles attitudes. We exploit ethnicity-level data on the number of slaves exported between 1400 and 1900 (i.e. before the Anglo-French division of Cameroon) to assess whether ethnicities located on the British side of the border were differently affected by the slave trade compared to groups on the French side. If this was the case, part of our results could be explained by such pre-colonial difference. We construct a measure of exposure to the trans-Atlantic slave trade following Nunn and Wantchekon [2011], and we merge it to our sample via latitude and longitude. Column (5) of Table 1.1 reports the coefficient obtained by estimating Equation 1.1 with the following outcome variable: the natural logarithm of one plus slave exports normalized by ethnic land area. We do not find evidence of a significant discontinuity in exposure to the trans-Atlantic slave trade for the two regions around the Anglo-French border.

Next, we run the same balance checks using ethnicity-level characteristics measured relying on respondents' self-reported ethnicity, rather than on their geographic location. While there is still no variation in plough use among respondents' ancestral societies, we can now additionally include balance checks for polygamy, brideprice, and dependence on gathering. Columns (6) to (11) show no clear discontinuity in any of these dimensions.

The second identifying assumption is no selective sorting around the treatment threshold. This would be violated if, for example, the Anglo-French division of Cameroon induced violence-prone individuals or ethnic groups to migrate from the French to the British part or *vice versa*. In this case, Cameroon's colonial past might still influence contemporary levels of IPV, but migration would be a main channel of persistence. That said, there is little evidence for selective migration from French to British Cameroon or *vice versa*. First, historical accounts suggest that the Anglo-French border acted as physical barrier to migration. In a letter about the first United Nations visit to British and French Cameroon in 1950, this was described as follows:

> *"A certain number of complaints were submitted to the Mission concerning the difficulties caused to the populations by the existence of a frontier and customs barriers between the two Cameroons. [...] The Mission heard more specific complaints re-*

> *garding [...] the impossibility, for certain families, clans and tribes, of maintaining normal relations between their members settled on both sides of the frontier. [...] Some say that the partition of the Cameroons into two zones is an arbitrary measure taken without the consent of the people and which denies them the relations they should have with the inhabitants of the neighboring territories."*— United Nations Trusteeship Council [1950b]

Second, in Figure A.3, we provide some evidence on the extent of selective migration by ethnic groups from and to areas close to the historical colonial border.[8] To that end, we merge ethnic groups in the Ethnographic Atlas with self-reported ethnic affiliations in the DHS (following the procedure suggested in Alesina *et al.* [2021]), and aggregate the data at the cluster level by assigning the ethnicity of the majority group to each cluster.[9] Based on this, we compare contemporaneous ethnic settlement patterns to pre-colonial ones by projecting the location of ethnic clusters in the DHS onto the Ethnographic Atlas. This descriptive exercise reveals that ethnic settlements have remained fairly stable over time and that there has been no systematic movement of entire ethnic groups across the border, in neither direction.

1.5 Long-Run Effects on Women's Outcomes

1.5.1 Participation in the Labor Market

The divergent opportunities that opened up for women when Cameroon was split into a French and a British colony may have had a long-term impact on female participation in activities outside the household, especially in the labor market. Thus, we now assess whether there is a discontinuity in today's female employment at the historical Anglo-French border. Table 1.3 reports estimates of Equation 1.1, for women located within 40 kilometers of the former Anglo-French colonial boundary. We display results using a naive OLS regression without the RD polynomial in Column (1), a local linear polynomial in distance to border in Column (2), a local linear polynomial in latitude and longitude in Column (3), and local linear polynomials in both in Column (4). All specifications control for survey year fixed effects, border segment fixed effects, distance to the capital, and distance to Douala.

In Panel A of Table 1.2, we estimate Equation 1.1 using as dependent variable a dummy indicating whether a woman is employed or not. Across all for specifications, we find evidence of a statistically significant and quantitatively important long-run colonizer effect on the probability that partnered women engage in economic activities. Specifically, women in former British territories are 16 percentage points more likely to be employed than their counterparts in former French territories. This estimate, which remains stable across the different specifications of the RD polynomial, is large in magnitude: it compares with a mean female employment rate of 73%.

[8] We use ethnicity to conduct this exercise because DHS does not collect information on individuals' migration history.

[9] We are successful for 41% of the 40km-sample clusters, and for 9 ethnic groups in the Atlas.

Table 1.2: *Effect of British Colonization versus French Colonization on Women's Economic Empowerment*

		Linear polynomial in:		
	OLS	Distance to the boundary	Latitude and longitude	Dist., lat. and long.
	(1)	(2)	(3)	(4)
Panel A: Female Employment				
British ($\hat{\gamma}$)	0.169***	0.161***	0.170***	0.161***
	(0.052)	(0.013)	(0.007)	(0.013)
Mean Dependent Variable	0.730	0.730	0.730	0.730
Panel B: Female Paid Employment				
British ($\hat{\gamma}$)	0.233***	0.240***	0.232***	0.240***
	(0.058)	(0.054)	(0.053)	(0.055)
Mean Dependent Variable	0.622	0.622	0.622	0.622
Observations	2,030	2,030	2,030	2,030
Clusters	358	358	358	358
Year fixed effect	Yes	Yes	Yes	Yes
Border fixed effects	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes

Notes: OLS estimates of Equation (1.1) for a window of 40 kilometers around the boundary. Standard errors clustered by DHS survey cluster are reported in parentheses. Panel A reports coefficients for female employment; Panel B reports coefficients for female paid employment. In Panels A-C 36.6% of observations are located on the British side. Geographic controls include distance to the capital and distance to Douala. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

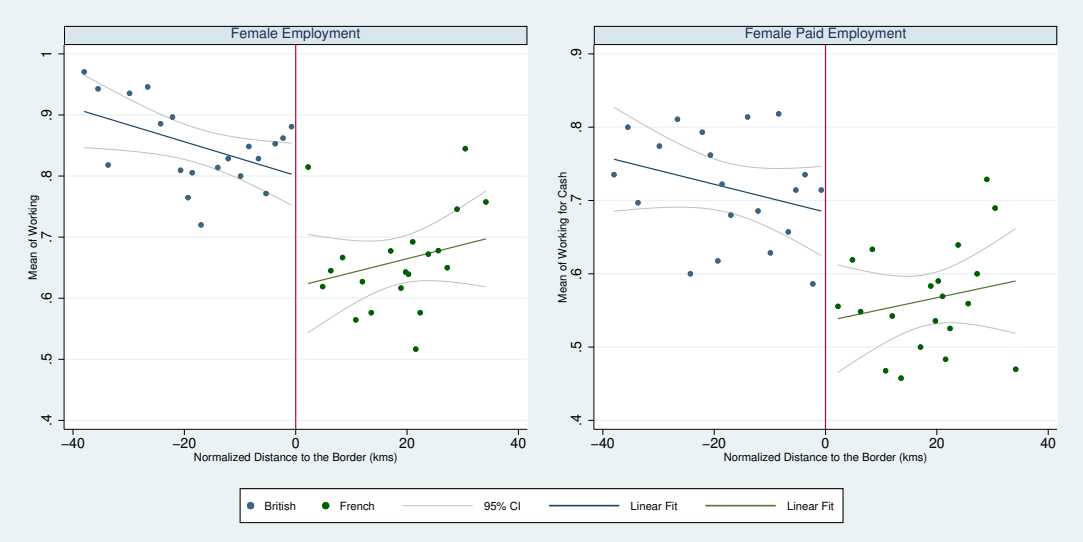
In the context of sub-Saharan Africa, it is important to distinguish whether individuals work for cash, are being paid in kind or not paid at all. The results in Panel B of Table 1.2 show that women in former British territories are over 38% more likely than their formerly French counterparts to be paid in cash.[10] Figure 1.4 provides the graphical analogue to the results above; it shows standard two-dimensional RD plots using distance to the border as the running variable.[11]

Appendix Table A.1 shows the results of several robustness checks. Columns (1) to (5) allow for different specifications of the RD polynomial. As can be seen, the estimated coefficient remains stable and significant for an interacted local linear polynomial in distance [Column (1)], and for the following higher orders of the RD polynomial: quadratic in distance [Column (2)]

[10] Consistent with the British colonizer effect on female paid employment, we find that women on the British side of the border are also more empowered in terms of control over household resources and property (See Appendix Table A.4).

[11] In Table A.7, we show that the results for female employment are not driven by specific sectors, such as agriculture, manual occupations, or services.

Figure 1.4: *Employment: RD Plots*



and latitude-longitude [Column (4)], and cubic in distance [Column (3)] and latitude-longitude [Column (5)].[12]

Column (6) shows estimates resulting from weighting the observations using a triangular kernel—assigning higher weight to individuals located very close to the boundary—and using an interacted linear polynomial in distance to the boundary. For both outcomes of interest, coefficients are significant and similar in magnitude compared that those from our baseline specification. Column (7) shows the results of a Donut exercise, which excludes observations within 5 kilometers to the boundary. This robustness check is important to verify that the results are not driven by systematic differences between border populations and populations residing further away from it. We find that all results remain robust. The coefficients remain significant and stable also when excluding clusters located in Douala, the economic capital of Cameroon [Column (8)].

Although we have shown that regions on the two sides of the border display balanced geographic and ethnic baseline characteristics (see Table 1.1), we nevertheless control for these characteristics in an alternative specification. Controlling for altitude, climate, a dummy for urban clusters, ancestral dependence on fishing, endogamous practices, and exposure to slave trade increases the magnitude of our estimates, as Column (9) shows. We reach the same conclusion by adding these controls one-by-one, or by controlling separately for geographic features and ancestral characteristics. Column (10) shows that our results are robust to additionally controlling for ethnicity fixed effects.[13] We consider this to be a powerful robustness check, as it allows us to account for ethnic-specific characteristics. Moreover, the mere fact that we can estimate the RD with ethnicity fixed effects means that individuals belonging to the same ethnic

[12] When denoting latitude by y and longitude by x, a quadratic RD polynomial takes the form: $x+y+x^2+y^2+xy$. A cubic polynomial in latitude and longitude equals: $x+y+x^2+y^2+x^3+y^3+x^2y+xy^2$.

[13] For this exercise, we use women's self-reported ethnicity in the DHS. In our sample, there are 38 different ethnic groups

group live on both sides of the border, and this reinforces the idea that the historical border split ethnic groups.[14] In column (11), we control for border segment fixed effects interacted with wave fixed effects. Again, estimates remain virtually unchanged.

Finally, columns (12) to (15) show the results of falsification checks which shift the border by 40 kilometers, both west and east. Irrespective of whether we adopt distance to the border or latitude-longitude as the running variable, this exercise does not reveal any placebo-boundary effect for neither of the two measures of female employment.

Appendix Figure A.4 probes the robustness of our estimates to different choices of bandwidth, controlling for linear and quadratic polynomials in latitude-longitude, distance to the boundary or both. The graphs depict coefficients and relative confidence intervals resulting from estimating Equation 1.1 for samples falling into various windows around the border, by 1-kilometer increments. Overall, coefficients remain relatively stable in magnitude and significance for bandwidths ranging between 10 and 100 kilometers.[15]

1.5.2 Domestic Violence

Given the British colonizer effect on female (paid) employment, and given that female employment has been linked to domestic violence in numerous settings (Anderberg *et al.* 2016; Cools and Kotsadam 2017; Bhalotra *et al.* 2018), Cameroon's colonial past might continue to influence contemporary levels of IPV. Thus, in Table 1.3, we re-estimate Equation 1.1 using two measures of domestic violence. In Panel A, we use as dependent variable a dummy indicating whether a woman has been exposed to physical and/or sexual violence in the year preceding the survey. In Panel B, the dependent variable is the sum of different types of physical and/or sexual aggression to which a woman has been exposed.

Across all four specifications in Panel A, we find that women on the British side of the historical border face a 10-11 percentage points higher risk of (past year) spousal violence than their counterparts on the French side. This is effect is not only precisely estimated but also large in magnitude; it compares with a mean prevalence of IPV of 28% throughout the region examined. The findings in Panel B show that women on the British side of the border are also exposed to higher-intensity IPV than those of on the French side: throughout the region examined, women report, on average, 0.58 different types of physical and/or sexual aggression by their partner; for those located on the British side of the border, this intensity index is more than 50% higher. Figure 1.5 illustrates the IPV results graphically, showing standard two-dimensional RD plots using distance to the border as the running variable. Taken together, the results suggest that the two colonial regimes in Cameroon continue to exert an influence on intra-household outcomes decades after their disappearance.

[14] As an alternative to adding ethnicity fixed effects, in non-reported regressions we re-run the same specifications with religion fixed effects. Results remain unchanged.

[15] Former British regions extend up to approximately 100 kilometers from the colonial boundary. The latter window is thus the natural upper bound for this exercise.

Table 1.3: *Effect of British Colonization versus French Colonization on IPV*

		Linear polynomial in:		
	OLS	Distance to the boundary	Latitude and longitude	Dist., lat. and long.
	(1)	(2)	(3)	(4)
Panel A: Any physical violence last year				
British ($\hat{\gamma}$)	0.113***	0.116***	0.101**	0.102**
	(0.043)	(0.043)	(0.044)	(0.043)
Mean Dependent Variable	0.283	0.283	0.283	0.283
Panel B: Number of different violent acts				
British ($\hat{\gamma}$)	0.319***	0.326***	0.290**	0.295***
	(0.109)	(0.107)	(0.112)	(0.111)
Mean Dependent Variable	0.576	0.576	0.576	0.576
Observations	2,030	2,030	2,030	2,030
Clusters	358	358	358	358
Year fixed effect	Yes	Yes	Yes	Yes
Border fixed effects	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes

Notes: OLS estimates of Equation (1.1) for a window of 40 kilometers around the boundary. Standard errors clustered by DHS survey cluster are reported in parentheses. Panel A reports coefficients for the binary variable; Panel B reports coefficients for a violence index ranging between 0 and 6. 36.6% of observations are located on the British side in all columns. Geographic controls include distance to the capital and distance to Douala. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

Figure 1.5: *IPV: RD Plots*

Our estimates for IPV are highly robust to the choice of bandwidth (see Appendix Figure A.5). Moreover, they pass all the robustness and placebo checks that we have conducted for female employment outcomes (see Appendix Table A.2).

1.5.3 Linking the Discontinuities for Female Employment and Domestic Violence

A natural question to ask next is whether the two separate border effects on female paid employment and IPV are explained by the same subsample of women, i.e., are those women who are more likely to work for cash also those who experience an increased likelihood of domestic violence? Or are the two border effects explained by two different subsamples of women, with one subsample driving the border effect on employment and the other the driving the border effect on IPV? The latter is a possibility as the partition of Cameroon and its later reunification induced a multiplicity of treatments. As a consequence, although women in formerly British areas face better employment opportunities than those in French areas, the incidence of IPV may be uniformly higher in British areas irrespective of whether women hold paid jobs or not. As an example, this could occur if the British rule, besides impacting women's employment opportunities, saw the introduction of a cultural norm that violence against women is acceptable.

To investigate these questions, let us denote by $V \in \{0,1\}$ whether a woman is IPV victim, by $E \in \{0,1\}$ whether she is in paid employment, and by $B \in \{0,1\}$ past French versus British colonial rule. Given that our data provides individual-level information on both V and E, we can decompose the probability of being IPV victim, $P(V)$, into the probability of being victimized and employed, $P(V \cap E)$, plus the probability of being victimized and not employed, $P(V \cap \not{E})$. Then, using the RD model, we estimate how each of these joint probabilities is affected by British colonial rule. Table 1.4 shows that the entire British colonizer effect on IPV (reported in Table 1.3) is driven by an increase in women's joint probability of being victimized and employed. By contrast, women's joint probability of being victimized and not employed is, if anything, higher in former French areas. Figure 1.6 shows two-dimensional RD plots corresponding to the just-described results.

Taken together, we conclude that the British colonizer effects on female paid employment and on IPV are not driven by two different subsamples of women. Instead, the only explanation consistent with our estimates is that those women who, due the legacy of British colonial rule, are more likely to be in paid employment are also those women who face an increased likelihood of being victims of IPV.[16]

[16] We have defined $P(V) = P(V \cap E) + P(V \cap \not{E})$, and estimated that $dP(V)/dB > 0$ is entirely driven by $dP(V \cap E)/dB > 0$, while $dP(V \cap \not{E})/dB \approx 0$ (see Table 1.4); the fact that $dP(V \cap E)/dB > 0$ implies that a switch from $B = 0$ to $B = 1$ causes some women to change victimization and/or employment status, and there are five possibilities for change: (a) from $\{V, E\} = \{0, 1\}$ to $\{V, E\} = \{1, 1\}$, (b) from $\{V, E\} = \{1, 0\}$ to $\{V, E\} = \{1, 1\}$, (c) from $\{V, E\} = \{0, 0\}$ to $\{V, E\} = \{1, 0\}$, (d) from $\{V, E\} = \{0, 0\}$ to $\{V, E\} = \{0, 1\}$, and (e) from $\{V, E\} = \{0, 0\}$ to $\{V, E\} = \{1, 1\}$. However, the changes of type (a) and (c) are inconsistent with $dP(E)/dB > 0$, while changes of type (b) and (d) are inconsistent with $dP(V)/dB > 0$; the only change consistent with $dP(V)/dB > 0$, $dP(E)/dB > 0$, and $dP(V \cap E)/dB > 0$ is that of type (e). Of course, the argument centers around net effects: it does not imply that there are no women who experience changes of types (a) to (d), but these changes would need to be offset by even more women who experience change of type (e).

Table 1.4: *Linking the Discontinuities for IPV and Paid Employment*

	$P(V \cap E)$			$P(V \cap \not{E})$		
	Dist	Lat-Lon	Dist Lat-Lon	Dist	Lat-Lon	Dist Lat-Lon
	(1)	(2)	(3)	(4)	(5)	(6)
British ($\hat{\gamma}$)	0.137***	0.124***	0.125***	-0.021	-0.023	-0.022
	(0.034)	(0.036)	(0.035)	(0.033)	(0.033)	(0.033)
Mean Dependent Variable	0.175	0.175	0.175	0.108	0.108	0.108
Observations	2,030	2,030	2,030	2,030	2,030	2,030
Clusters	358	358	358	358	358	358
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Border fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimates of Equation (1.1) for a window of 40 kilometers around the boundary. Standard errors clustered by DHS survey cluster are reported in parentheses. $P(V \cap E)$ denotes the joint probability of a woman being victimized and in paid employment. 168 out of 743 women (22.6%) are both victimized and employed on the British side; 187 out of 1,287 women (14.5%) are both victimized and employed on the French side. $P(V \cap \not{E})$ denotes the joint probability of a woman being victimized and not in paid employment. 73 out of 743 women (9.8%) are victimized and not employed on the British side; 146 out of 1,287 women (11.3%) are victimized and not employed on the French side. 36.6% of observations are located on the British side in all columns. Geographic controls include distance to the capital and distance to Douala. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

1.6 Mechanisms

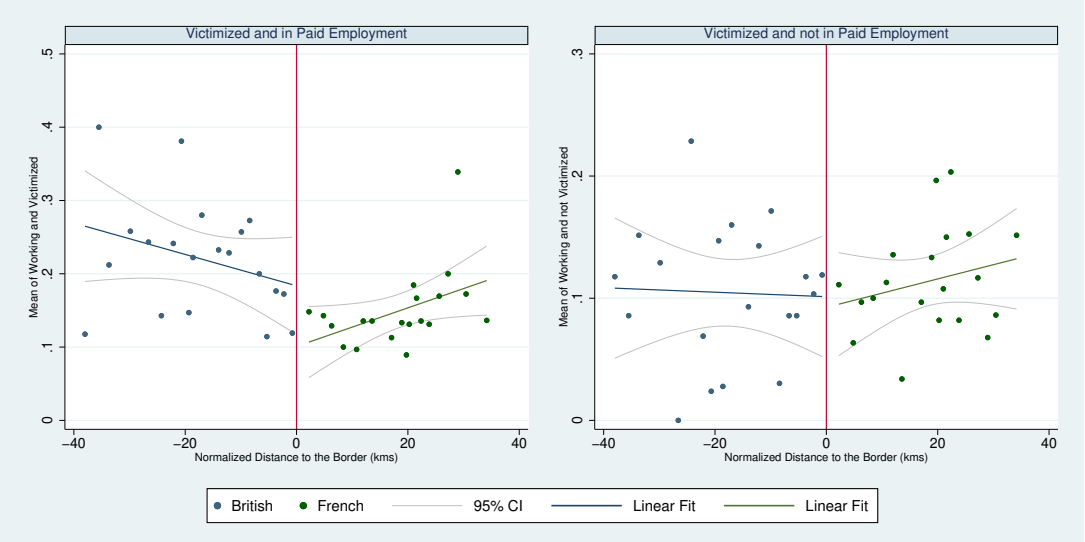
In this section, we explore possible explanations for the result that the British colonial rule caused a persistent increase in both female employment and IPV, with the two effects being driven by the same subsample of women.

1.6.1 Male Backlash?

To begin with, one class of economic models clearly inconsistent with our findings are theories of household bargaining that incorporate domestic violence (Aizer 2010; Anderberg *et al.* 2016). These models predict that higher female employment comes with a reduced risk of falling victim to IPV, as it levels the intra-household balance of bargaining power between partners. Yet, the border discontinuities we have identified show that female empowerment in terms of access to paid employment and a high vulnerability to IPV go hand in hand.

At first sight, this finding is consistent with the "male backlash" hypothesis, which is based on the idea that: better employment opportunities due British colonial ⇒ women work more ⇒ intra-household power relations are upend ⇒ male backlash in form of IPV. However, the results are also consistent with what we label the "resource extraction" hypothesis, whereby the partners of women who hold paid jobs use violence instrumentally to extract resources from them (Anderberg and Rainer [2013]). In this case: better employment opportunities ⇒ women work more ⇒ incomes available to them increase ⇒ instrumental violence to extract resources.

Figure 1.6: *Victimization and Employment: RD Plots*



A related possibility is what we call the "forced labor" hypothesis, which says that husbands use violence to force their partners to work more. In this case: better employment opportunities ⇒ instrumental violence to force female labor ⇒ women work more. Both the resource extraction and the forced labor hypothesis implicitly assume that men have a preference for high spousal incomes, as they both portray domestic violence as a means of maximizing men's own disposable income. By contrast, the male backlash hypothesis assumes that men do not want their partners to work, as it would threaten their own dominant role within the partnership.

In order to provide more direct evidence consistent with the notion of male backlash in our context, we exploit the fact that domestic violence is often accompanied by controlling and coercive behavior on the part of a partner. Crucially for our purpose, rich information in the DHS allows to consider two types of coercive control, namely work-related and non-work-related. Our idea is this: if male backlash is the explanation for the border discontinuity for IPV, we would expect it to be explained by domestic violence that occurs in connection with work-related coercive behavior on the part of a partner. By contrast, since domestic violence among women outside paid employment occurs, if anything, more frequently in formerly French areas, we would expect no pronounced border discontinuity for IPV that arises in connection with non-work-related coercive behavior.

To explore this idea, we first create a dummy variable for controlling behavior, C, which equals one if a woman reports at least one of the following seven types of partner behavior: (i) husband is jealous if she talks to other men; (ii) husband accuses her of infidelity; (iii) husband does not permit her to meet her girl friends; (iv) husband tries to limit her contact with family; (v) husband insists on knowing where she is; (vi) husband doesn't trust her with money; and (vii) husband does not want her to work/to have a job. Second, we create a dummy for work-related coercive control, W, which equals one if a woman answers item (vii) in the positive, i.e., reports that her husband does not want her to work or to have a job. Finally, we break up women's

probability of being IPV victim, $P(V)$, into the probability of being victimized and with a partner that shows work-related coercive control, $P(V \cap C \cap W)$, plus the probability of being victimized and with a partner who shows non-work related coercive control, $P(V \cap C \cap \not{W})$, plus the probability of being victimized and with a partner who shows no signs of coercive control, $P(V \cap \not{C})$.

In Columns (1) to (6) of Table 1.5, we first estimate the British colonizer effect on women's probability of having a partner who shows (i) work related coercive behavior and (ii) non-work related coercive behavior. We find that women on the British side of the border have a 10 to 12 percentage points higher propensity of having a partner of type (i) (at a mean of 17 percent). By contrast, women's propensity to have a partner of type (ii) is 6 percentage points higher (at mean of 69 percent) on the French side of border, although this estimate is statistically not significant. In Columns (7) to (15), we use the RD model to estimate how the three joint probabilities defined above are affected by past British colonial rule. Strikingly, we find the entire colonizer effect on IPV is explained by an increase in women's joint probability of being victimized and with a partner who objects their employment, while women's joint probability of being IPV victim and subject to non-work-related coercive partner behavior is as high in formerly French areas as it is in British territories. Decomposing the former joint probability effect further, we find it is almost entirely driven by an increase in the joint probability of being victimized, with a partner who objects to female employment, and in paid employment (see Appendix Table A.3).[17]

We view this evidence as consistent with the male backlash hypothesis, but incompatible with the resource extraction and the forced labor hypothesis. To be clear on the latter point: consistent with the resource extraction and forced labor hypothesis, and in line with the British colonizer effect on female paid employment, our data indeed reveals that women in former British areas belong to households that display a higher wealth status (see Appendix Table A.5).[18] But the fact that men on the British side of border object to their partners' employment speaks against higher IPV levels being driven by resource extraction or forced labor motives, since both center around a stronger male preference for high spousal earned incomes.

Finally, the results in this section are also in line with the following interpretation. Exposure to IPV may induce women to strive for financial independence in order to improve their divorce opportunities (i.e. their outside options from an abusive marriage), and this would lead to higher employment rates. In this case, compared to the male backlash hypothesis, causation runs from IPV to female employment and thus in the opposite direction. However, this interpretation seems implausible in our setting for two reasons. First, the rates of formal and informal separation in Cameroon are rather low. Only 1.2% of Cameroonian women aged 15-49 report being divorced at the time of the 2004 and 2011 DHS survey.[19] Since women can exit a

[17] Appendix Tables A.6 and A.7 provide the graphical RD analogues to these results.

[18] This conclusion is valid when considering only the sample of women selected for the DHS domestic violence module, the sample of all women, the sample of all men, or the sample of all women and men combined.

[19] The prevalence of divorce is even smaller for the sample of women within 40 km of the historical border (0.9%). In the sample of ever-partnered women who were interviewed for the domestic violence module, 1.4% are divorced.

Table 1.5: *Linking the Discontinuities for IPV and Husbands' Controlling Behavior*

	Controlling Behavior (including work-related)			Controlling Behavior (excluding work-related)			$P(V \cap C \cap W)$			$P(V \cap C \cap \not{W})$			$P(V \cap \not{C})$		
	Dist	Lat-Lon	Dist Lat-Lon	Dist	Lat-Lon	Dist Lat-Lon	Dist	Lat-Lon	Dist Lat-Lon	Dist	Lat-Lon	Dist Lat-Lon	Dist	Lat-Lon	Dist Lat-Lon
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
British ($\hat{\gamma}$)	0.102***	0.116***	0.116***	-0.056	-0.063	-0.063	0.074***	0.076***	0.077***	0.016	0.003	0.003	-0.000	-0.008	-0.008
	(0.034)	(0.034)	(0.034)	(0.045)	(0.045)	(0.045)	(0.023)	(0.025)	(0.025)	(0.035)	(0.036)	(0.036)	(0.012)	(0.012)	(0.012)
Mean	0.167	0.167	0.167	0.692	0.692	0.692	0.080	0.080	0.080	0.186	0.186	0.186	0.020	0.020	0.020
Observations	1,897	1,897	1,897	1,897	1,897	1,897	1,897	1,897	1,897	1,897	1,897	1,897	1,897	1,897	1,897
Clusters	357	357	357	357	357	357	357	357	357	357	357	357	357	357	357
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Border fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimates of Equation (1.1) for a window of 40 kilometers around the boundary. Standard errors clustered by DHS survey cluster are reported in parentheses. *Controlling Behavior (including work-related)* is a dummy equal to 1 if a woman reports controlling behaviors from the husband (jealous if talking with other men, accuses her of unfaithfulness, does not permit her to meet her girls friends, tries to limit her contact with family, insists on knowing where she is, doesn't trust her with money) including that the husband does not want her to work/have a job. *Controlling Behavior (excluding work-related)* is a dummy equal to 1 if a woman reports any controlling behavior from the husband, but not that the he does not want her to work/have a job. $P(V \cap C \cap W)$ denotes the joint probability of a woman being victimized and affected by her husband's controlling behaviors, including him opposing her employment. 68 out of 692 women (9.8%) are victimized and affected by their husband's controlling behaviors (including work-related) on the British side; 79 out of 1,205 women (6.6%) are victimized and affected by their husband's controlling behaviors (including work-related) on the French side. $P(V \cap C \cap \not{W})$ denotes the joint probability of a woman being victimized and affected by controlling behavior, but without the husband opposing to her employment. $P(V \cap \not{C})$ denotes the joint probability of a woman being victimized and not affected by controlling behaviors. Geographic controls include distance to the capital and distance to Douala. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

relationship even without divorcing, and since some women might not be formally married, we also consider the prevalence of informal separation. Only 4.6% of Cameroonian women report not living together with their partners at the time of the 2004 and 2011 DHS survey.[20]

Second, if higher female employment rates on the British side were a consequence of women's attempts to increase their outside options, then one would expect divorce or informal separation to be more common in former British territories. Instead, we find no statistically significant discontinuity in the prevalence of divorce or informal separation between former British and French regions. In addition, we do not find systematic differences in other aspects of relationships such as the prevalence of informal unions or partnership duration (see Table A.6).

1.6.2 Other Factors Affecting Female Employment and Domestic Violence

The evidence presented so far is suggestive that the British colonial rule triggered an increase in female employment, and that this in turn led to male backlash in the form of increased domestic violence. However, the British colonization might have had an impact on third factors, some observed and some unobserved, affecting female employment, IPV, or both. Take, just as an example, a high level of male alcohol abuse, which might require women to work more (e.g., to compensate for lower partner earnings) and put them at a higher risk of IPV. Thus, we now explore observed factors of this type and ask whether the historical Anglo-French boundary still exerts an influence on them. To this end, we exploit rich information in the DHS on women interviewed for the domestic violence module, their partners, and their households. In addition, we use geo-referenced UCDP GED conflict data.

In Tables 1.6 and 1.7, we re-estimate Equation 1.1 using as dependent variables the following: male employment, men's occupations, male alcohol consumption, education, education gaps, number of children, age, age gaps, practice of polygamy, and exposure to conflict. To preview our results, none of these other factors display discontinuities at historical Anglo-French border.[21]

First, we examine outcomes related to women's partners. Contrary to the results on female employment highlighted in the previous section, Panel A in Table 1.6 shows that male employment varies smoothly at the historical Anglo-French border.[22] Despite this, differential colonial labor market policies might have favored the development of different sectors on the two sides of the border. Men on the British side might have specialized in different occupations, and may have more stressful jobs that make them more violent at home. In addition, men's occupations might also directly affect whether women are employed or not. In Panels B, C, and D we show that there is no discontinuity in whether men work in agriculture, in manual jobs, or in services,

[20] Among women within 40 km of the historical border, the prevalence is 4.5%. Among ever-partnered women in our sample, 5.3% are not living with their partner.

[21] These conclusions remain unchanged when computing Romano-Wolf stepdown p-values, based on standard errors adjusted for multiple hypothesis testing.

[22] This measure of male employment naturally refers to a specific subset of the male population, i.e. male partners of eligible women interviewed for the domestic violence module: this explains the large percentage of employed men in the sample.

Table 1.6: *Other Factors: Partner's Characteristics*

	A. Male Employment			B. Men Agriculture			C. Men Manual		
	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
British ($\hat{\gamma}$)	-0.002	0.002	-0.003	-0.021	0.033	-0.022	-0.007	-0.042	-0.006
	(0.013)	(0.007)	(0.013)	(0.066)	(0.059)	(0.064)	(0.058)	(0.053)	(0.058)
R-W p-value	0.614	0.683	0.604	1	0.871	0.990	1	0.762	1
Mean	0.976	0.976	0.976	0.291	0.291	0.291	0.351	0.351	0.351
Observations	2,016	2,016	2,016	2,016	2,016	2,016	2,016	2,016	2,016
Clusters	358	358	358	358	358	358	358	358	358
	D. Men Services			E. Male Alcohol Consumption			F. Men's Education		
	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
British ($\hat{\gamma}$)	-0.004	-0.011	-0.004	-0.036	-0.044	-0.034	0.566	0.454	0.600
	(0.028)	(0.027)	(0.028)	(0.045)	(0.046)	(0.045)	(0.473)	(0.448)	(0.449)
R-W p-value	1	0.931	1	0.802	0.713	0.851	0.535	0.673	0.426
Mean	0.090	0.090	0.090	0.167	0.167	0.167	8.413	8.413	8.413
Observations	2,016	2,016	2,016	1,417	1,417	1,417	1,940	1,940	1,940
Clusters	358	358	358	348	348	348	358	358	358

Notes: OLS estimates of Equation (1.1) for different outcomes for a sample within 40 kilometers from the boundary. Each column controls for a different specification of the RD polynomial. All regressions include geographic controls (distance to the capital and distance to Douala), together with border segment and year fixed effects. *Male Employment*: equals 1 if the husband is employed, 0 otherwise; *Men Agriculture*: 1 if husband is employed in the agricultural sector, 0 otherwise; *Men Manual*: 1 if husband is employed in a manual occupation, either skilled or unskilled, 0 otherwise; *Men Services*: 1 if husband is employed in services, 0 otherwise; *Male Alcohol Consumption*: 1 if husband gets drunk often, 0 otherwise; *Men's Education*: husband's years of education. R-W stands for Romano-Wolf p-values adjusted for multiple testing. Standard errors clustered by DHS survey cluster are reported in parentheses. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

the three largest employment sectors in our sample.[23] Taken together, this suggests that men's employment is unlikely to explain our main results on women's employment or IPV.

Many studies across different disciplines have acknowledged a close association between male alcohol abuse and violent behavior.[24] Using Mexican data, Angelucci [2008] has shown that a cash transfer increasing women's income led to a reduction in their partners' alcohol abuse and to lower levels of IPV. Men's alcohol abuse might also be directly linked to women's employment, if women have to compensate for lower partner earnings. Thus, we now ask whether former

[23] We repeat this exercise with all sectors proposed by the DHS classification: (i) professors, technicians, or managers; (ii) clericals; (iii) sales; (iv) agriculture (self-employed); (v) agriculture (employee); (vi) household and domestic; (vii) services; (viii) skilled manual; (ix) unskilled manual. We find a significant positive discontinuity only for sector (i), and only in one out of our three preferred specifications. Results of this exercise are available upon request.

[24] A review of the evidence from psychiatry and criminology can be found in Leonard [2005].

French and British areas differ in the extent to which men abusively consume alcohol. Panel E shows that husbands in former British areas do not display a higher propensity of alcohol abuse than their formerly French counterparts.[25]

Another potential determinant of IPV is women's and men's education (Erten and Keskin [2018]), which determines their earnings potential. This, in turn, has possible implications for who's got the say in household decisions and, therefore, for IPV (Hidrobo *et al.* [2016]). As mentioned above, education is one of the spheres in which British and French colonial practices had very different implications, particularly for women: those in British territories benefited from a protestant education system that aimed at equally educating girls and boys, while French policies centered around educating a small administrative elite. Our point estimates in Panel F of Table 1.6 and Panel A of Table 1.7 suggest that, throughout the region examined, both men and women on the British side of the historical border have, on average, more years of education than their French counterparts. However, these coefficients are not statistically significant, with the exception of one for women's education.[26] In Panel B of Table 1.7, we show that the education gap between partners is not a channel explaining higher IPV or female employment on the British side either, as it varies smoothly at the boundary.

In Panel C of Table 1.7, we examine fertility. The reproductive role of women is crucial in traditional African societies.[27] Therefore, this dimension has to be taken into account in any attempt to explain IPV on the African continent. In particular, if women have a preference for fewer children than their male partners, and an improved intra-household bargaining position causes this preference to be more strongly reflected in couples' actual fertility choices, then this may lead to a male backlash effect. Lower fertility could also explain higher female participation in the labor market. The coefficients in Panel G indicate that fertility is unlikely to be a main channel for the IPV and employment effects we have uncovered: the estimates are relatively small in magnitude and far from being statistically significant.[28]

Next, we consider two other potential correlates of IPV: women's age and spousal age gaps. While women's age varies smoothly at the border (Panel D), we find a marginally significant effect on spousal age gaps, with women in former British regions being more likely to be partnered with older men on average, compared to women in former French regions. Yet, when applying the Romano-Wolf multiple hypothesis correction, these estimates lose significance at conventional levels, suggesting that they are unlikely to explain our main findings.

[25]Note that he outcome variable in Table 1.6 takes the value one when women report their husband getting drunk often. We also obtain insignificant estimates when allowing the dependent variable to mirror the frequency of drinking episodes, taking either value 0 (never), 1 (only sometimes) or 2 (often).

[26]Note that the coefficient in Panel A, column (3), becomes insignificant once accounting for multiple testing. To explore other measures of women's educational attainment, in Table A.8 we use an indicator for whether women completed primary school, and find a positive and significant effect. We do not find a similar effect on women's probability of having completed secondary school, and, perhaps surprisingly, we find a reversed effect on a measure of literacy. However, this estimate is only significant at the ten percent level.

[27]It has been widely explored in the work of Lesthaeghe [1989], according to whom "The reproductive function itself is so crucial [...] that the status of adulthood for women is almost completely contingent on motherhood and the last installments of bridewealth payments are often transferred upon the birth of the first child only."

[28]The outcome variable referred to in Table 1.7 is the number of living children. However, the same conclusion holds when considering the number of children ever born.

Table 1.7: *Other Factors: Education, Fertility, Age, Spousal Imbalances, Conflict*

	A. Women's Education			B. Partners' Education Gap			C. Number of Children		
	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
British ($\hat{\gamma}$)	0.611	0.398	0.623*	0.025	0.096	0.034	0.044	0.138	0.044
	(0.383)	(0.361)	(0.372)	(0.330)	(0.305)	(0.325)	(0.211)	(0.200)	(0.211)
R-W p-value	0.228	0.614	0.168	1	0.931	1	1	0.792	1
Mean	7.221	7.221	7.221	1.160	1.160	1.160	2.934	2.934	2.934
Observations	2,030	2,030	2,030	1,940	1,940	1,940	2,030	2,030	2,030
Clusters	358	358	358	358	358	358	358	358	358

	D. Women's Age			E. Spousal Age Gap			F. Polygamy			G. Exposure to Conflict		
	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
British ($\hat{\gamma}$)	0.204	0.195	0.206	1.397*	1.456*	1.385*	0.035	0.050	0.034	0.061	-0.023	0.064
	(0.782)	(0.779)	(0.781)	(0.834)	(0.778)	(0.803)	(0.039)	(0.036)	(0.038)	(0.112)	(0.103)	(0.102)
R-W p-value	1	0.931	1	0.168	0.119	0.139	0.792	0.347	0.802	0.614	0.871	0.455
Mean	31.33	31.33	31.33	8.291	8.291	8.291	0.165	0.165	0.165	0.145	0.145	0.145
Observations	2,030	2,030	2,030	1,824	1,824	1,824	1,749	1,749	1,749	2,030	2,030	2,030
Clusters	358	358	358	358	358	358	356	356	356	81	81	81

Notes: OLS estimates of Equation (1.1) for different outcomes for a sample within 40 kilometers from the boundary. Each column controls for a different specification of the RD polynomial. All regressions include geographic controls (distance to the capital and distance to Douala), together with border segment and year fixed effects. *Women's Education*: woman's years of education; *Partners' Education Gap*: husband's years of education minus wife's years of education; *Number of Children*: number of living children; *Women's Age*: respondent's age in years; *Spousal Age Gap*: husband's age minus wife's age; *Polygamy*: 1 if woman is in a polygamous union; *Exposure to Conflict*: 1 if there was any conflict event in the woman's arrondissement of residence since 1980. R-W stands for Romano-Wolf p-values adjusted for multiple testing. Standard errors clustered by DHS survey cluster are reported in parentheses. Columns (19) to (21) report standard errors clustered at the *arrondissement* level. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

The position and the role of women in society is inextricably linked to marriage structure, which may determine both the risk of IPV and women's participation in economic activities outside the household (Tur-Prats [2017], Koenig *et al.* [2003]). One particular case in point is polygamy, which is widespread in sub-Saharan Africa and in Cameroon, and has been shown to affect important economic outcomes (Tertilt [2005]). Moreover, the prevalence of polygamy was reduced by colonial practices in general and by missionary education in particular (Fenske [2015]). In light of this evidence, we examine whether British versus French colonization had a long-lasting impact on the prevalence of polygamy, which according to Murdock's Ethnographic Atlas was customary among the groups inhabiting the geographical area we study (Figure A.2). The results displayed in Panel F show that 16.5% of all interviewed women are in a polygamous union, and that the prevalence of polygamy is somewhat higher on the British side of the boundary. However, the estimated border discontinuities are not statistically significant at conventional levels.

Exposure to armed conflict in the early years of life has important consequences in adulthood. As far as domestic violence is concerned, La Mattina and Shemyakina [2017] show how sub-Saharan African women who experienced conflict in their childhood are more likely to report and accept abuse perpetrated by their partners, due to reduced educational attainment. Civil conflict might also lower women's decision making via a change in the marriage market's sex ratio. In post-genocide Rwanda, this lead to increased spousal abuse (La Mattina [2017]). Changes in sex ratios can also impact women's labor market participation. Thus, we now probe whether women on the British side of the historical border have witnessed more conflict events as compared to their French counterparts. This exercise draws upon geo-referenced UCDP GED conflict data. We construct a binary measure of conflict exposure, which assesses whether there was any conflict event in the woman's *arrondissement*[29] of residence since 1980. Panel G shows that the regions on the two sides of the former Anglo-French boundary were relatively peaceful, with an average of only 0.15 conflict events. Moreover, we do not detect a border discontinuity in this variable.[30] We conclude that the colonial division did not have a long-lasting impact on conflict, and that the latter is an unlikely driver of the British colonizer effect on IPV.

1.7 Self-Reporting of IPV

A concern with the interpretation our results might be that the self-reporting of IPV is affected by women's empowerment and therefore by different colonial practices in the past. We believe this not to be a caveat. As mentioned at the outset, the DHS uses rigorous ethical and privacy protocols which avoid generic and subjective questions such as "Have you ever experienced domestic violence?" and instead employs questions about specific episodes of violence. This

[29] *Arrondissements* are the lowest administrative units in Cameroon.

[30] In non-reported regression, we also used a fatality count measure as dependent variable, and found no evidence of a border discontinuity. For this specification, we cluster our standard errors at the district level, but we obtain similarly insignificant results when clustering at the *arrondissement* level. Moreover, we conducted the same exercise using DHS clusters as units of observation rather than individuals. Again, no significant effect is recorded.

ameliorates the concern that less-empowered women may not identify what they are subject to as violence since they consider it as the norm. Despite this, respondents may still perceive an exposure risk when asked directly about IPV episodes. Since the disutility costs associated with this are likely heterogeneous, truthful reporting of IPV may differ across more and less empowerment women. However, the direction of such reporting bias, if any, is far from clear. On one side, it could be argued that empowering women changes their attitudes to gender equality and, hence, increases the likelihood of reporting IPV in their daily lives. However, it is equally possible that empowered women—i.e., those who do not fit the typical victim stereotype—face larger disutility of being exposed as victims and therefore underreport IPV. Agüero and Frisancho [2017] use an experimental approach based on indirect questioning techniques to assess the extent of truthful IPV-reporting in the DHS. They find no evidence of IPV-misreporting for low-educated women, but high-educated women appear to underreport IPV victimization. This suggests that, if anything, empowered women underreport IPV rather than being more daring to report it.

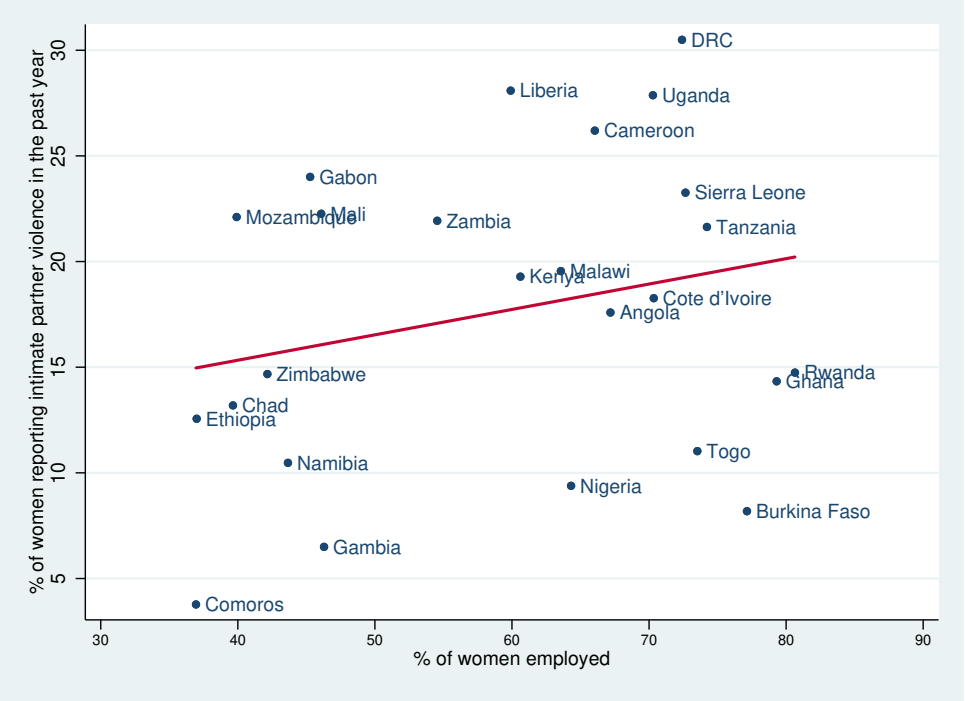
Building on this finding, if IPV-reporting of empowered women is a threat to our interpretation of the results, we would expect a positive individual-level correlation between IPV and different measures of female empowerment. Multivariate regressions based on our estimation sample show this not to be the case. As could be expected given the results in Section 1.6.1, we find a positive correlation between a woman's experience of IPV and paid employment. However, there is a negative correlation between domestic violence and a woman's education. The latter correlation is inconsistent with the notion that IPV-reporting increases with a woman's empowerment.

1.8 Conclusion

Although there exist many empirical investigations of the importance of Africa's colonization for subsequent economic development, much less is known about its role in shaping the lives of women. In this paper, we have made an attempt to fill this gap by utilizing a natural experiment of history: the partition of Cameroon into a British and a French colony. The two colonial regimes opened up divergent economic opportunities for women in an otherwise culturally and geographically homogeneous setting. In particular, women in British territories gained opportunities to earn cash wages under the same conditions as their male counterparts. In contrast, the French colonial practice invested in the male employment-dominated infrastructure sector. Thus, Cameroon's colonial past offers a unique opportunity to investigate how policies promoting female empowerment affect women's outcomes in the long-term.

Our main finding shows that the British colonial rule has two legacies that are still visible today. On the one hand, it empowered women economically in terms of access to employment, being paid in cash wages, and control of household resources. On the other hand, it made women highly vulnerable to IPV. This result is incompatible with household bargaining models that incorporate domestic violence but it is consistent with theories of male backlash.

Figure 1.7: *Female Employment and IPV in sub-Saharan Africa*



Notes: The figure collapses individual-level data on IPV and employment for a sample of women aged 15-49 from all sub-Saharan countries surveyed by DHS.

There are some broader implications of our results. It has been widely recognized for some time that women's economic empowerment and development are closely related. Some striking examples of the positives of empowerment are highlighted in the World Bank's 2012 World Development Report: equalizing access to productive resources between female and male farmers could increase agricultural output in developing countries by 2.5-4%, and eliminating barriers that prevent women from working in certain sectors or occupations could increase labor productivity by as much as 25% in some countries.

What has been much less emphasized is that women's economic empowerment is unlikely to come without unwanted consequences. In particular, increased access to jobs and financial resources can upend intra-household power relations and instigate backlash in the form of IPV. Descriptive evidence from many parts of the developing world shows that there is indeed a strong positive correlation between women's employment and domestic violence. The situation in sub-Saharan Africa, inhabited by over 10% of the world's population, is a particular case in point. Figure 1.7, based on data from the Demographic and Health Surveys, shows a strong positive correlation between female employment and IPV across all surveyed countries in sub-Saharan Africa. Although nothing causal can be read into this correlation pattern, it is hard to reconcile with the notion that female economic empowerment, here access to jobs, translates into less violence against women.

From a policy perspective, our results should not be read as challenging the much-discussed and well-justified case for women's economic empowerment. But there is a cautionary tale here. In our context, institutions that offered women new opportunities in the labor market brought

with them substantial risk in the form of increased IPV. Thus, paradoxically, empowerment in the labor market and disempowerment at the micro-level of the family appear to be two sides of the same coin. To ameliorate this by-product of female economic empowerment, enforceable laws that offer women direct legal protection from domestic violence and the opportunity to divorce from abusive partners would seem of paramount importance.

Chapter 2

Cultural Distance and Conflict-Related Sexual Violence*

2.1 Introduction

Ethnic civil conflict accounts for roughly half of all civil conflicts around the world during the period 1946-2005 (Wimmer *et al.* [2009]). While the relationship between ethnicity and conflict has received considerable attention by the scholarly literature, we lack knowledge on what specific dimensions of ethnicity bring about conflict. Ethnic groups differ in many aspects, like religion, language, gender, or social structure. However, it is unclear which of these ethnic cleavages become salient in a given violent situation, and how they determine the way combatants fight. In this paper we introduce, for the first time, distance in ethnic-specific gender norms as a potential trigger of sexual violence in conflict, a largely understudied technology of war.

Sexual violence in armed conflict is one of the most brutal forms of violence against women. It is a widespread crime that encompasses, among others, acts of rape, sexual slavery, and forced prostitution (International Criminal Court [2002]). At least 500,000 women were raped during the Rwandan genocide (April-July 1994), 50,000 during the Bosnian war (1992-1995), 250,000 during the Sierra Leonean civil war (1991-2002), 200,000 in the Bangladesh liberation war (1971), and 400,000 in a single year of the ongoing conflict in Eastern Congo (Meger [2016]). This phenomenon comes with disastrous long-lasting physical and psychological consequences for victims, their families, and their communities (Ba and Bhopal [2017]).

Despite being widespread, armed-related sexual violence is not ubiquitous; its prevalence and intensity vary considerably both *across* and *within* conflicts (Skjelsbaek [2001]). Why do some actors systematically rape while others never do so? We propose and test two main hypotheses for explaining the occurrence and intensity of sexual violence in armed conflicts in Africa:[1] (1) gender-unequal ethnic actors are more likely to perpetrate sexual violence, and (2)

*This chapter is based on joint work with Ana Tur-Prats.

[1]We restrict our analysis to Africa for two reasons. First, the ethnographic information is better documented and systematized for this continent. Second, the concordance table that we use to merge ethnic groups from the

sexual violence increases with the cultural distance in gender norms between the two actors. Additionally, we test whether cultural distance in gender norms can explain the use of general (not sexual) violence, and whether other measures of cultural distance—linguistic or religious—determine the use of conflict-related sexual violence.

To this end, we build on the Sexual Violence in Armed Conflict (SVAC) dataset (Cohen and Nordås [2014]), which comprises all civil conflicts between 1989 and 2009 and includes an index of the intensity of sexual violence that ranges from 0 (no sexual violence) to 3 (massive and systematic sexual violence). We combine this dataset with other sources to include information on the ethnic identity of both actors involved and their ancestral socioeconomic characteristics. More precisely, we use the dyadic version of the SVAC dataset, GEO-SVAC (Bahgat *et al.* [2016]), to add information on the identity of both actors involved in the conflict (i.e., government/state military and rebel forces). We then assign to each actor its ethnic identity using the Ethnic Power Relations (EPR) dataset (Vogt *et al.* [2015]). Finally, we use the Murdock Ethnographic Atlas to add information on the ethnic characteristics of each actor. The resulting dataset has a dyadic and bidirectional structure and contains information on the intensity of sexual violence as well as on the ethnic characteristics at the actor-conflict-country-year level.[2] Our dataset covers 128 actors (106 related pairs of actors) involved in 33 ethnic civil conflicts fought in 27 different countries spanned over the period 1989-2009 (N=623).

The empirical investigation includes several steps. We first generate and validate an ethnic gender inequality index (thereafter, eGII) and then use this index to test our two main hypotheses. To construct the eGII, we start by identifying nine ethnic traits that, according to well-established interdisciplinary literature, relate to anthropological notions of gender (in)equality:[3] matrilineality, patrilocality, stem family types, dependence on gathering, hunting, agriculture, husbandry, pastoralism, and the use of the plough. Next, we use principal component analysis (PCA) to construct an Africa-wide gender-inequality index for each ethnicity based on these nine ethnic characteristics. Consistently with anthropological notions, the first component loads positively on ancestral arrangements conducive to gender-unequal norms—e.g., patrilocality and dependence on male-dominated activities like animal husbandry and pastoralism—and negatively on ancestral traits conducive to gender equality like matrilineality and dependence on agriculture, in which women played a prominent role.

We then show that our eGII is a powerful predictor of an ethnic group's gender norms elicited through the Afrobarometer and the Demographic and Health Survey. The eGII tallies well with

EPR dataset to the Murdock Ethnographic Atlas, provided by Michalopoulos and Papaioannou [2016], covers only African countries.

[2]For example, for the conflict that in 1994 confronted the Government of Chad against the *Comité de Sursaut National pour la Paix et la Démocratie* (CSNPD), we observe the prevalence and intensity of sexual violence that the Government of Chad exerted against the CSNPD (intensity 2) and the sexual violence that the CSNPD exerted against the Government of Chad (zero). At the same time, we also observe the ethnicity of both actors: the rulers of the Government of Chad at that time came from the Zaghawa and Bideyat ethnic groups, while the CSNPD rebel forces were formed by Sara soldiers. Table B.2 illustrates this.

[3]We mostly draw from the social-sciences literature in economics (e.g., Boserup [1970], Alesina *et al.* [2013], Lowes [2017], Becker [2018], Tur-Prats [2019]), and anthropology (e.g., Schneider and Gough [1961], Sanday [1973], Friedl [1978], Korotayev [2003]).

contemporary measures of gender inequality such as female employment, attitudes towards gender and towards wife beating, and the sexual component of intimate partner violence. Our eGII offers two advantages compared with these contemporary measures of gender inequality. First, since the eGII is based on ancestral characteristics and captures deep-rooted gender norms, it is plausibly unaffected by contemporaneous institutions and recent conflict-history. Second, it summarizes a broad range of domains related to gender inequality into a single dimension, which facilitates cross-cultural studies of gender norms. Additionally, we find a large within-country variation in the eGII, which speaks in favor of the use of ethnic-specific gender inequality indexes as opposed to the existing country-wide indexes, especially in ethnically diverse regions like Africa.

Having validated the eGII, we then test our first hypothesis—namely, that gender-unequal actors are more likely to perpetrate sexual violence—and indeed find that the eGII is positively associated with sexual violence in conflict. This association holds when we include conflict fixed effects, year fixed effects, conflict-specific time trends, and victim's ethnic characteristics.[4] This result is in line with previous findings on gender-based violence during peacetime (Alesina *et al.* [2021]; Tur-Prats [2019]) but had not been previously tested in the context of war.

In a next step of our empirical analysis, we test our second hypothesis, i.e., whether sexual violence in conflict increases with the cultural distance in gender norms between the combatants. We exploit the dyadic structure of the data, and run a specification in the spirit of a gravity equation, similarly to recent literature adjusting the canonical trade models to other contexts (Becker *et al.* [2020]; Grosjean [2011]; Serafinelli and Tabellini [2017]; Spolaore and Wacziarg [2009]). By regressing the intensity of sexual violence on the absolute distance in gender norms between perpetrator and victim, we find that the larger the cultural distance between the ethnic belligerents, the higher the intensity of sexual violence. As is standard in this literature, we include a battery of controls (conflict and year fixed effects, and conflict-specific time trends) to net the effect of potential confounders.

When examining this association further, we uncover that it is driven by a specific *cultural clash*: sexual violence increases when the perpetrator holds more gender-unequal norms than the victim, but not viceversa. This result remains significant when we isolate the effect of cultural distance from the combatant's own characteristics by separately including perpetrator and victim fixed effects. This allows us to rule out potential alternative explanations for our results that have to do with actors' time-invariant factors correlated with their gender norms. For example, we can rule out that the effect is driven by the aggressive nature of more gender-unequal perpetrators, by more gender-equal victims being more likely to report sexual violence, or that sexual violence is used strategically to harm women in gender-equal societies, where women represent a valuable asset. These results are also robust to a comprehensive battery of robustness tests, such as the inclusion of additional fixed effects (conflict-year, country), to

[4]Continuing with our previous example, the Government of Chad in 1994 had an eGII value of 0.52, whereas the rebel group CSNPD had an eGII value of 0.28.

alternative versions of the eGII, and to abstracting from the temporal variation in the data, among others.

We offer a new identity-based explanation for these findings (Akerlof and Kranton [2000]). This explanation rests upon the notion of cultural distance, and is consistent with the asymmetry of our results, i.e., the fact that we only find positive effects of cultural distance on sexual violence when perpetrators are *more* gender-unequal than their victims, but not viceversa. When confronting a more gender-equal society, perpetrators might perceive the relatively better position of women as a threat to their own norms, and resort to sexual violence to alleviate the negative feelings experienced by this encounter. Conversely, combatants that encounter a more gender-unequal society might not experience any menace or necessity to react against the different role of men and women in the opponent's society.

Cultural distance in gender norms could be correlated with other dissimilarities between ethnic groups, which in turn could trigger violence in general, and not only sexual violence. To test this, we conduct a placebo exercise in which we re-run our analysis replacing sexual violence with a measure of general violence: the number of deaths inflicted by the perpetrator on the victim.[5] Our results show that cultural distance in gender norms does not explain general violence within a conflict.

Finally, we explore whether general cultural differences, and not only differences in gender norms, can explain sexual violence in conflict. We exploit two widely-used measures of cultural distance: linguistic and religious distance. Following Fearon [2003], we construct a measure of linguistic distance for each related pair of actors. Albeit positively correlated with distance in gender norms, we do not find that linguistic distance is associated with sexual violence.

Similarly, religious distance between combatants is not powerful in explaining the use of sexual violence. Controlling for religious distance in our preferred specification leaves the main coefficients unchanged and, if at all, religious distance between perpetrator and victim is negatively associated with sexual violence.[6] Taken together, these results suggest that conflict-related sexual violence is not driven by general cultural differences, and its understanding requires a gender-based explanation. Our eGII allows us to capture the specific dimension of cultural distance that matters to explain this widespread behavior.

This paper contributes to several strands of literature. First, we contribute to the interdisciplinary literature on sexual violence in armed conflict—summarized in the next Section—by advancing and empirically testing a new hypothesis, namely that the prevalence and the intensity of war-related sexual violence is explained by a *clash of conceptions* on what is the appropriate role of men and women in society. Second, this study adds to the literature on gender norms and gender inequality by proposing and validating a new gender inequality index based on ethnic traits, and enriches the growing literature on the long-run cultural determinants

[5]Because this measure is also bidirectional, we can replicate our analysis replacing sexual violence by general violence as our dependent variable.

[6]These findings are robust to alternative specifications in which we isolate the component of cultural distance that is unexplained by differences in gender norms. We first regress linguistic or religious distance on distance in gender norms, and use the residuals of this regression as our main explanatory variable.

of violence against women (Alesina *et al.* [2021], Tur-Prats [2019]) by empirically investigating a so-far overlooked form of gender-based violence. Third, our paper is related to the literature on how ancestral conditions, by persistently shaping cultural norms, can influence contemporary outcomes and behaviors (Alesina *et al.* [2013]; Becker [2018]; Guiso *et al.* [2016]; Voigtländer and Voth [2012]).

Finally, we add to work on ethnic conflict (see Blattman and Miguel [2010] for a summary) both on the empirical and conceptual front. From an empirical standpoint, we construct a novel dyadic dataset that includes information on the ethnic identity and ancestral characteristics of *all* actors involved in an ethnic civil conflict. Our conceptual contribution stems from the acknowledgment that previous work has focused on understanding whether ethnic diversity *triggers* war, but has remained silent on the role that ethnicity plays in shaping the technology of war. By analyzing the intensive margin of violence, we hypothesize and show that cultural distance in gender norms between actors can explain how violence *unfolds* once conflict takes place.

The remainder of the paper is structured as follows. Section 2 reviews the related literature and advances our hypotheses. Section 3 describes the data and the procedure used to merge the various sources, and presents some descriptive statistics. In Section 4 we construct and validate our eGII. Section 5 tests our first hypothesis, namely the impact of the perpetrator's gender norms on the use of sexual violence, and provides robustness tests. In Section 6 we test the cultural clash hypothesis, run a battery of robustness checks, and explore the relationship between gender-norms distance and general violence as well as the relationship between other linguistic and religious distance and sexual violence in conflict. Finally, Section 7 concludes.

2.2 Existing Literature and Hypotheses

The first hypothesis we test is whether gender-unequal ethnic actors are more likely to perpetrate sexual violence during a conflict. The idea that gender norms in a society and the prevalence of gender-based violence are linked has been advocated by scholars across various disciplines. Two contributions in the economics literature have empirically shown that deeply entrenched norms about the role of men and women in society are associated to intimate partner violence (IPV). In Africa, women belonging to ethnic groups where, in ancestral times, women had more marginalized roles in the economy and society relative to men are today more likely to be IPV victims (Alesina *et al.* [2021]). In Spain, Tur-Prats [2019] finds that historical family structures—stem versus nuclear—influenced women's participation in non-domestic work and persistently shaped gender roles. Areas where stem families were predominant in the past are characterized by more progressive gender norms, and women residing in these regions are today less likely to report IPV.

Among scholars investigating the determinants of gender-based violence in the context of armed conflict, some favored a socio-cultural rationale behind soldiers' use of sexual violence. Through the analysis of previous literature, Skjelsbaek [2001] noticed a consensus in consid-

ering sexual violence a weapon of war, i.e., part of a pre-meditated strategy, rather than the manifestation of a latent biological need triggered by a state of war.[7] In Skjelsbaek's [2001] conceptualization, perpetrators use sexual violence strategically to empower (i.e., masculinize) their own identity and to victimize (i.e., feminize) the opponent's. According to this view, any attempt to analyze sexual violence in conflict without considering gender relations is incomplete. Meger [2016] also lists context-specific gender norms—in turn shaped by political, economic, and social structures—as one of the factors underlying the occurrence of conflict-related sexual violence. These views, in turn, are in line with anthropological research on sexual coercion. For example, Sanday [1981b] suggests that rape is an expression of cultural forces operating at the societal level. Through the analysis of a cross-cultural sample of tribal societies, she shows that rape-prone societies are characterized, among other things, by a higher degree of male dominance, compared to non rape-prone ones.

The first empirical analysis of the determinants of sexual violence in armed conflict was conducted by Cohen [2013], who finds support for the so-called *combatant socialization theory*, according to which soldiers recruited by force use rape as a method to socialize and generate cohesion. Contrary to the conjectures of the aforementioned literature, Cohen [2013] does not find a relationship between gender inequality and sexual violence in conflict at the country level. Albeit positive, the correlation between rape and fertility—used as a proxy for gender inequality—is not statistically significant. We test the gender inequality hypothesis by moving from the *country level* to the *conflict-actor's level*, and by measuring gender inequality through ethnic actors' deeply-rooted cultural norms, in turn shaped by their ancestral economic, societal, and family arrangements.

Our main hypothesis is that sexual violence is a function of cultural distance in gender norms between two opposing ethnic belligerents. To the best of our knowledge, there is no existing theory or empirical evidence on how cultural distance in gender norms between ethnic groups might influence their decision to perpetrate sexual violence during a conflict. However, anthropological work by Taylor [1999] emphasizes the gender component of the violence that unfolded during the Rwandan genocide, which materialized in systematic acts of sexual violence perpetrated by Hutu men against Tutsi women. According to Taylor [1999], this was the result of Hutu men disagreeing with Tutsi women's prominent role in society:

> *"Hutu extremists aimed at reclaiming the lost ground of patriarchy and re-asserting a male dominance that had probably never existed in Rwanda's actual history. [...] The Rwandan genocide was not simply a battle for political supremacy between groups of men, it was also about re-configuring gender. [...] Gender relations were falling into a state of decadence and disorder as more [Tutsi] women attained positions of prominence in economic and public life."*

[7] Inspired by early anthropological work by Symons [1979], a much-discussed contribution by Thornhill and Palmer [2001] describes rape through the lenses of evolutionary biology. According to the authors, rape is a biologically-determined behavior: it is either the direct result of an evolutionary adaptation to increase men's reproductive success, or a byproduct of other adaptive traits, such as aggressiveness.

Similar triggers are mentioned by Nowrojee [1996] in her report for Human Rights Watch about the Rwandan genocide:

> *"Tutsi women were targeted on the basis of the genocide propaganda which had portrayed them as calculated seductress-spies bent on dominating and undermining the Hutu. Tutsi women were also targeted because of the gender stereotype that portrayed them as beautiful and desirable, but inaccessible to Hutu men whom they allegedly looked down upon and were "too good" for."*

Boserup [1970] also describes the prominent role of Tutsi women, and mentions that these different social patterns were a source of extreme tensions between Hutu and Tutsi. Both tribes were characterized by a caste system, and besides both Hutu men and Hutu women worked as agricultural laborers for the Tutsi upper class. She notices that:

> *"The wives of the Tutsi chiefs had absolute power over most male members of the local communities, while the Hutu women were at the bottom of the social hierarchy, doing the hard labour and subordinate to all other groups in the communities, including their own husbands."*

In a very different context such as the Spanish Civil War (1936-1939) we also find evidence in line with our cultural distance hypothesis. Sexual violence was widespread during this conflict, and was particularly perpetrated by the rebel group against those women that remained loyal to the Republican government (Preston [2012]). The rebel fascists considered that these women had transgressed the traditional and patriarchal model they defended (Muñoz-Encinar [2019]).

Furthermore, our hypothesis is grounded in previous literature that analyzes how cultural distance between two entities can trigger a range of violent manifestations, from discrimination (see Becker's [1957] seminal work) to conflict. At the broad macro-cultural level, Huntington's [2000] 'clash of civilizations' thesis states that cultural and religious differences are the main determinants of conflict in the post World War II era. Focusing on interstate wars, Bremer [2000] argues that more ethnically distant societies will be more likely to fight against each other. Closer to our study, Caselli and Coleman's [2013] model of ethnic conflict predicts that ethnic groups are more likely to clash when the differences between them are more pronounced. Based on these theoretical insights, we hypothesize that cultural distance between ethnic groups may not only trigger conflict, but also influence in what ways violence unfolds once conflict takes place.[8]

[8]Empirical findings on the relationship between cultural distance and the onset of conflict are mixed, and suggest that the direction of this association may depend on the nature of conflicts. When looking at intrastate conflicts, Arbatli *et al.* [2013] find that genetically diverse countries are more likely to engage in civil war. Conversely, Spolaore and Wacziarg [2016a] show that genetic distance between any two countries is associated with less international conflict with each other.

2.3 Data

We construct a novel dataset, which combines a variety of sources on ethnic conflict, the actors involved, their use of war-related sexual violence, and their ancestral ethnic characteristics. This Section provides an overview of the main data sources used for the analysis. More details on the data sources as well as on the procedure we adopted to construct the dataset can be found at the Section B.1 in the Appendix.

2.3.1 Sexual Violence in Armed Conflict

The source of our dependent variable is the Sexual Violence in Armed Conflict (SVAC) Dataset (Cohen and Nordås [2014]), which contains information on sexual violence used in civil conflicts fought between 1989 and 2009. We focus on *ethnic* civil conflicts, which are defined as "armed conflicts between the government of a state and one or more internal opposition group(s) that cause at least 25 battle-related deaths within a year and in which armed groups (i) explicitly pursue ethno-nationalist aims, motivations, and interests; and (ii) recruit fighters and forge alliances on the basis of ethnic affiliation" (Gleditsch *et al.* [2002]). We exploit the *dyadic* version of the SVAC dataset, called GEO-SVAC (Bahgat *et al.* [2016]), which includes both the identity of the perpetrator of sexual violence and the identity of the other actor involved in the conflict. Consistently with the definition of civil conflict, one of these two actors is always the government of a state, and the opponent is always a rebel group.

Adhering to the International Criminal Court's rationale, SVAC defines war-related sexual violence as including acts of rape, sexual slavery, forced prostitution, forced pregnancy, forced sterilization, and forced abortion (International Criminal Court [2000]). In addition, following Wood [2009], sexual mutilation and sexual torture are also included. SVAC draws upon annual reports from three sources (Amnesty International, Human Rights Watch, and the US State Department) to construct a measure of prevalence of sexual violence at the conflict-actor-year level. The resulting variable is an index ranging between 0-3 that reflects the magnitude of the phenomenon. More specifically, it takes the value 3 if, in a given year of conflict, an actor perpetrated acts of massive, innumerable, or systematic sexual violence and if reported incidents or victims of sexual violence exceeded 1,000; 2 if sexual violence was described as widespread and common, and reports of victims or incidents ranged between 25 and 999; 1 if reported victims and incidents were below 25 and the occurrence of sexual violence was only isolated; 0 if no sexual violence was mentioned in a given year in relation to a specific conflict.[9]

2.3.2 Conflict Actors' Ethnic Identity and Ancestral Characteristics

To assign an ethnic identity to each conflict actor—rebel groups and governments—we exploit the rich information provided by the Ethnic Power Relations (EPR) Dataset Family (Vogt *et al.*

[9]For further details on the methodology of data collection and coding refer to Cohen and Nordås [2014].

[2015]). EPR defines an ethnic group as "an identity group that defines itself or is defined by others along linguistic, religious, or racial characteristics".

The EPR dataset family contains information, *inter alia*, on ethnic groups' involvement in civil war as part of a rebel organization and on ethnic groups' access to executive government power. We are therefore able to link each rebel force and each government to one or multiple EPR ethnic groups, depending on whether rebels and governments are the result of ethnic alliances.[10] Section B.1.4 in the Appendix illustrates this merging procedure with a concrete example of a conflict event in Liberia.

We add information on ethnic groups' ancestral characteristics using the Ethnographic Atlas (EA), coded by Murdock [1967] and updated by Nunn and Wantchekon [2011]. The EA is arguably the most compelling source of ethnographic information for 1,265 societies around the world, collected at the end of the 19th century. For Africa, the EA provides detailed information on groups' socio-economic conditions, settlement patterns, and family arrangements prior to European contact. We will describe these variables in detail in Section 4.

We add the information provided by the EA to the dataset on conflict through the concordance table provided by Michalopoulos and Papaioannou [2016], which links 196 EPR groups to 593 ethnicities in the EA using a variety of sources. In some cases, this matching procedure results in a one-to-one mapping between EPR groups and EA groups. For example, the ethnic group of the rebel force FLEC-FAC in Angola, the *Cabindan Mayombe*, is matched with the *Yombe* group in the EA. However, in other cases, a conflict actor is associated to multiple EA groups either because (i) an actor is represented by multiple EPR groups, (ii) an EPR group corresponds to multiple groups in the EA, or (iii) both.

An example of the latter case is the following: the RFDG rebel group in Guinea is composed of members belonging to the EPR groups called *Malinke* and *Peul*. In turn, the Michalopoulos and Papaioannou's [2016] correspondence table matches *Malinke* to four EA groups (*Yalunka, Konyanke, Malinke,* and *Koranko*), and the *Peul* to three EA groups (*Foutadjalon, Sokoto, Liptako*). In these instances, we weight the ethnic characteristics of each EPR group by the size of the EA groups to which it corresponds. In the just-mentioned example, *Peul*'s dependence on pastoralism will be a weighted average between *Foutadjalon's, Sokoto's*, and *Liptako's* dependence on pastoralism, based on the three ethnic groups' size, in turn proxied by the land area covered by their settlements. We will provide estimates using both the weighted and the unweighted version of the ethnic characteristics, and show that our results are generally insensitive to this procedure.

The final sample includes 33 ethnic civil conflicts fought between 1989 and 2009 in 27 African countries, involving 128 different actors (106 related pairs of actors). The resulting sample size is N=623.

[10] We assume that state and rebel military forces mirror the ethnic composition of governments and rebel groups, respectively. In Section 5.4 we conduct a robustness test in which we assume that state forces mirror the ethnic composition of the entire country (weighted by the size of each group's settlement area) and show that our results are robust to this alternative definition.

2.3.3 Descriptive Statistics

Sexual violence was present, in some level of intensity, in 82% of the conflicts included in our sample. 21% of all ethnic civil conflicts in Africa between 1989 and 2009 experienced at least one episode of sexual violence at the highest intensity, i.e., involving at least 1,000 victims. State forces perpetrate sexual violence more frequently than rebel groups. However, when perpetrated by rebel groups, the intensity of sexual violence is on average higher.

Figure 2.1 reports the spatial variation of sexual violence at the ethnicity level, conditional on the ethnic group being involved in a conflict. The striking picture that emerges is that there is considerable within-country variation in whether or not ethnic groups (and therefore, armed actors) engage in sexual violence. An interesting example is that of Algeria. In the long civil war between the government and various rebel armed forces, which began in 1991, rebel groups never made use of sexual violence, while the government constantly engaged in it throughout the war. In other cases, such as the one of Sudan, the vast majority of ethnic groups involved in conflict perpetrated sexual violence, but there was a quite large variation in the intensity of it. Finally, as in the case of Nigeria, the use of sexual violence was widespread across groups, and its incidence homogeneous.

2.4 Ethnic Traits and Gender Norms

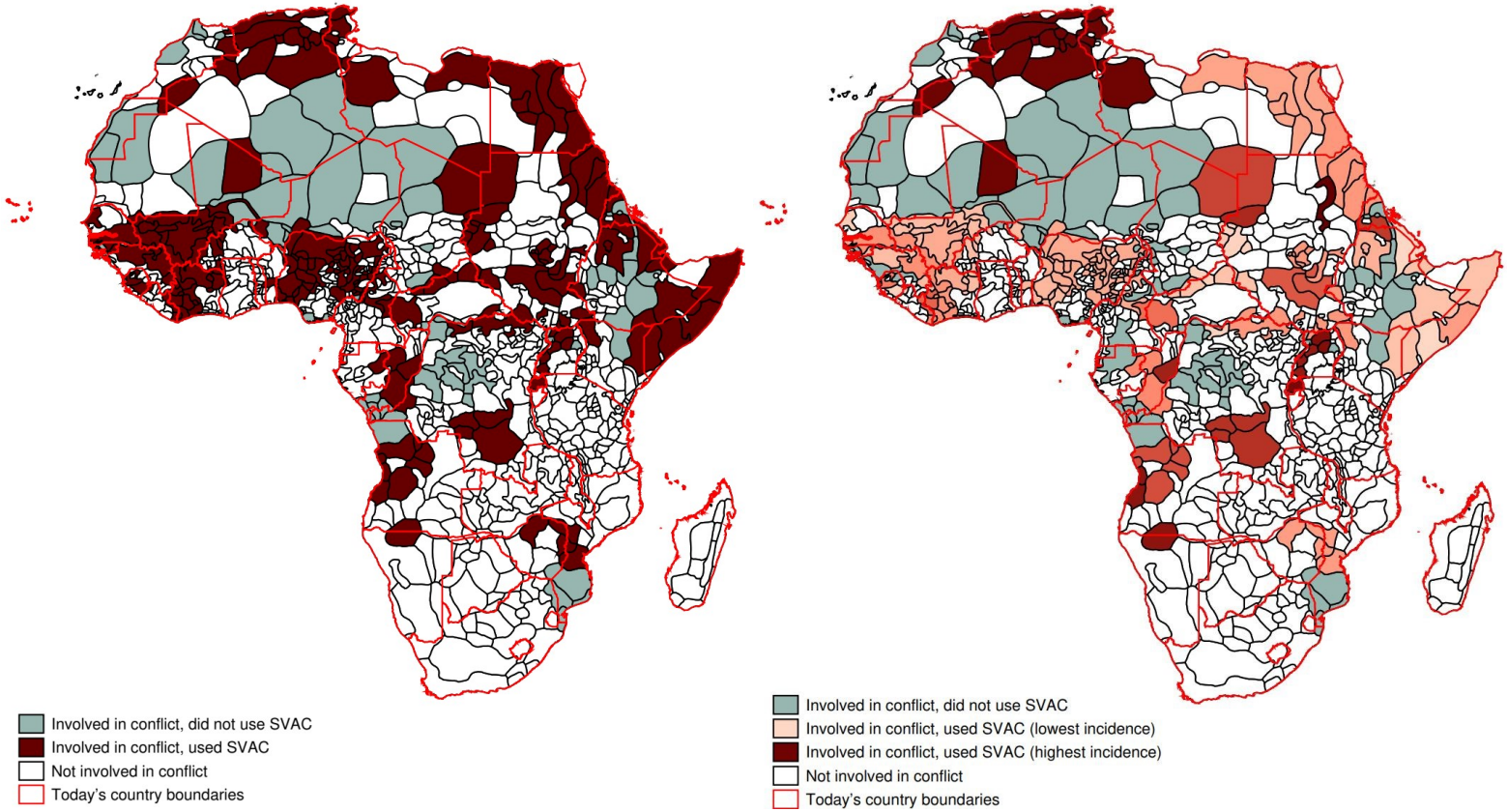
This section outlines the procedure we adopted for selecting and grouping ethnic characteristics into an Africa-wide ethnic gender-inequality index. Established literature spanning different disciplines has demonstrated that ancestral economic and societal arrangements have persistently shaped gender relations (see Giuliano [2018] for a review). Our choice of ethnic characteristics is informed by what this literature has highlighted as relevant determinants of gender norms in a society.[11]

2.4.1 Descent, Residence and Family Arrangements

Anthropologists have argued that societies where descent, residence, and family arrangements are centered around women tend to be characterized by higher gender equality (Martin and Voorhies [1975], Sanday [1981a]).

[11] We do not consider ethnic characteristics on which the literature is inconclusive. For example, it is unclear whether the practice of brideprice increases or decreases gender inequality. On the one hand, brideprice is a recognition of women's value, and it is more typical in societies where women have an important role in agricultural production (Boserup [1970]). On the other hand, the obligation of women to pay back the brideprice in case of divorce may decrease their bargaining power. The association between polygyny and gender equality is also ambiguous. On the one hand, women's status in polygynous unions may be lower, in particular for younger wives due to early marriage and large age gaps with the husband. Alesina *et al.* [2021] show that women in polygynous unions are more likely to suffer from intimate partner violence across Africa, but, at the same time, women and men in societies that traditionally practice polygamy are less likely to justify intimate partner violence. Furthermore, as highlighted in Boserup [1970], polygyny is more typical in societies where women constitute an important economic asset.

Figure 2.1: *Ethnic Groups' Use of Sexual Violence in Conflict (1989-2009) in Africa*



Notes: Left: involvement in conflict and use of sexual violence by Murdock ethnic groups in Africa. Right: involvement in conflict and incidence of sexual violence used by Murdock's ethnic groups in Africa, through an index varying between 0-1, which captures the total incidence of sexual violence in armed conflict for the period 1989-2009. Sources: Murdock Ethnographic Atlas and GEO-SVAC dataset.

In matrilineal societies, inheritance is traced through female family members. Therefore, women are key for determining descent and have constant support from their kin network (Schneider and Gough [1961]). Lowes [2017] has shown that women belonging to matrilineal ethnic groups in the Democratic Republic of Congo detain higher bargaining power within the household compared to their counterparts in patrilineal societies, and that they are less likely to be victims of intimate partner violence. As shown in Gneezy *et al.* [2009], matrilineal women are also more likely to display behavioral traits that are usually typical of men—such as willingness to compete—ones that have often been advocated as factors explaining economic disparities between men and women. Taken together, this evidence supports the notion that matrilineal societies, when compared to patrilineal ones, are characterized by more equitable gender norms.

Lineage systems and kinship structures in a society are inextricably linked to residence patterns. Patrilineal societies are also likely to be patrilocal, a system of postmarital residence where the newly formed couple moves near the husband's kin group (Murdock [1967]). In these societies, women may be less protected by their own family, and husbands may more easily exercise their authority over women. Scholars in anthropology have hypothesized that patrilocality is a direct consequence of women's low economic participation (Korotayev [2003]).

Tur-Prats [2019] has linked the prevalence of stem family types to higher gender equality. The co-residence of the wife with the mother-in-law frees up women from the burden of domestic work, and allows them to exercise a productive role in the economy and participate in family subsistence.

Based on this literature, we consider matrilineality, patrilineality, patrilocality, and stem family types as relevant ethnic traits capturing gender norms in a society.

2.4.2 Subsistence Activities

In ancestral societies, the relative participation of women and men in economic activities has persistently shaped gender relations (Friedl [1978], Sanday [1973]). According to Friedl [1978], in hunter-gatherer societies men exerted control over animal protein, a scarce and hard to acquire resource. Since hunting activities require a certain amount of physical strength, they are predominantly a men's task. As a result, these societies tend to be characterized by high degrees of male dominance. The same applies to societies whose subsistence is highly dependent on animal husbandry.

Pastoralism is a specific type of animal husbandry based on herd animals that require natural pasture, and entailed frequent and extended periods of male absence from the community, resulting in higher paternity uncertainty. Becker [2018] shows that these byproducts of pastoralism—male absence and paternity uncertainty—incentivized the adoption of measures to control women's sexuality and mobility. Women in societies where pastoralism was historically an important source of subsistence are today more likely to be infibulated, to be restricted in their mobility, and to hold more gender-unequal attitudes.

In the African context, where agriculture was characterized by shifting cultivation, female participation in agricultural activities was traditionally high, as emphasized by Boserup [1970] and confirmed in the ethnographic data by Murdock [1967]. In contrast, in other regions of the world where plough agriculture was more common, the traditional division of labor was reversed, with men taking up the majority of agricultural work and women remaining confined to the domestic sphere. This agricultural system based on plough agriculture—and the consequent division of labor—contributed to the evolution of gender-unequal norms, as empirically demonstrated in Alesina *et al.* [2013].

In light of this literature, we consider dependence on hunting, gathering, pastoralism, agriculture, and the use of the plough, as relevant ethnic characteristics to describe gender norms in a society.

2.4.3 Ethnic Gender Inequality Index: Construction and Validation

We use the just-described nine ethnic characteristics to construct an Africa-wide index for each ethnic group using principal component analysis (PCA). Information on each characteristic comes from the Murdock Ethnographic Atlas. Three of these nine traits—matrilineality, stem family, and dependence on agriculture—are reconcilable with notions of gender equality. Therefore, we expect these single traits to be *negatively* correlated with the eGII. The remaining six ethnic characteristics—the use of the plough, patrilocality, dependence on gathering and hunting, dependence on pastoralism, and animal husbandry—have been associated with gender inequality. In turn, we expect these single traits to be *positively* correlated with the eGII.

The first principal component alone explains 32% of the common variance of the nine variables across Africa. Table 2.1 presents the loadings of each of the nine traits included, i.e., their correlation with the first principal component. The sign of the loadings is as expected for the majority of ethnic traits. Matrilineality and dependence on agriculture are negatively correlated with the first component, while patrilocality, dependence on pastoralism, the use of the plough, and dependence on animal husbandry are positively correlated. Instead, the correlation between stem families and the first component is particularly small, and with an unexpected sign. Similarly, dependence on hunting and gathering are negatively correlated with the fist principal component, despite what is argued in the anthropological literature.[12]

Due to the presence of these "ambiguous" ethnic traits, we also provide an alternative version of the eGII, i.e., a restricted one based exclusively on ethnic characteristics that are unambiguously linked to gender (in)equality.[13] Table B.10 in the Appendix reports the corresponding loadings for these five ethnic traits, which hold the expected sign. When discussing our main results, we will show that they are robust to this alternative version of the eGII.

We normalize the predicted score of the PCA to range between 0 and 1, with 0 denoting highest gender equality and 1 denoting highest gender inequality. Figures 2.2 and 2.3 report

[12] Actually, Alesina *et al.* [2021] find that descendants of societies that depended more heavily on hunting are less prone to justifying intimate-partner violence today.

[13] We exclude the dependence on gathering and hunting, stem family, and the use of the plough.

Table 2.1: *eGII: PCA loadings*

Variables	*Loading*
Gender Equal Traits	
Matrilineal	-0.26
Dependence on agriculture	-0.27
Gender Unequal Traits	
Virilocal	0.30
Dependence on pastoralism	0.55
Use of the plough	0.29
Dependence on animal husbandry	0.55
Ambiguous Traits	
Stem	0.01
Dependence on gathering	-0.15
Dependence on hunting	-0.26
Kaiser-Meyer-Olkin's measure of sampling adequacy	0.58

Notes: Loadings from the principal component analysis on the eGII.

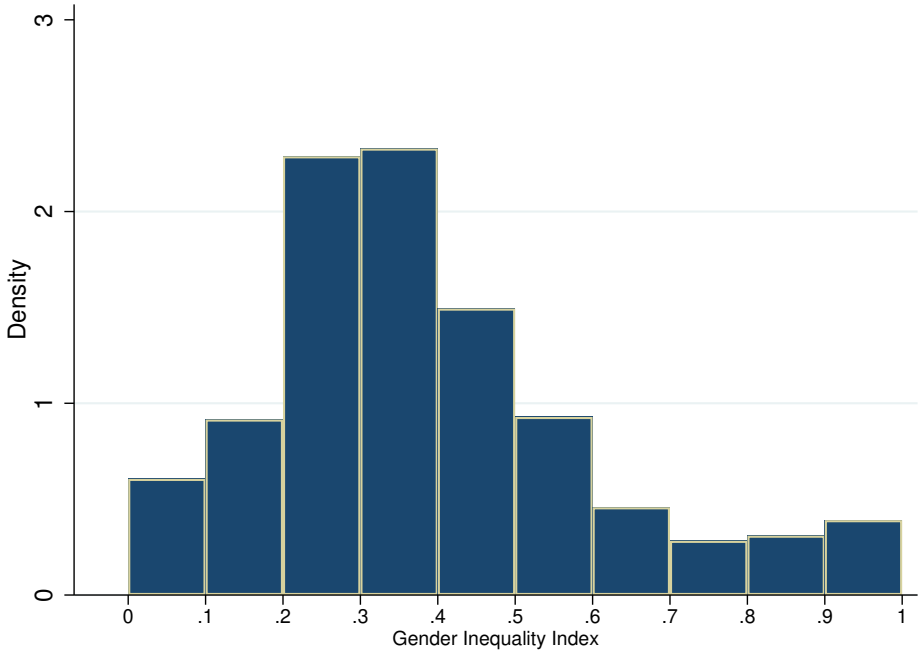
the distribution of the eGII across Africa and in our sample, respectively. Ethnic groups in our sample, i.e. those that were involved in at least one conflict between 1989 and 2009, tend to be characterized by more gender unequal norms on average, compared to the average of the continent as a whole.

Figure 2.4 reports the geographical distribution of the eGII across Africa, displaying the Murdock ethnic map and the corresponding eGII for each group. The highest levels of gender inequality are prevalent in ethnic groups located in North and East Africa, while the lowest are concentrated in Central Africa and in some parts of West Africa. The distribution of the eGII varies considerably across regions, but also within countries. One of the most extreme cases is Tanzania, where ethnic groups span from the lowest bin of the eGII (0-0.25) to the highest bin (0.75-1). Figures B.5 and B.4 in the Appendix display the distribution of the restricted version of the index, which is very similar to the one of the main eGII. Finally, figure 2.5 compares the distribution of the eGII with the distribution of sexual violence in armed conflict.

Our eGII is correlated with proxies for gender (in)equality today. Figure 2.4 shows that countries with the lowest rates of female labor force participation are also those in which ethnic groups are characterized by a high eGII. We further validate our eGII by using micro-data from the Demographic and Health Survey (DHS) and the Afrobarometer survey. We match individuals' self-reported ethnicity to the Murdock Atlas via the Linking Ethnic Data from Africa (LEDA) algorithm (Müller-Crepon *et al.* [Forthcoming]).[14] The advantage of this micro-

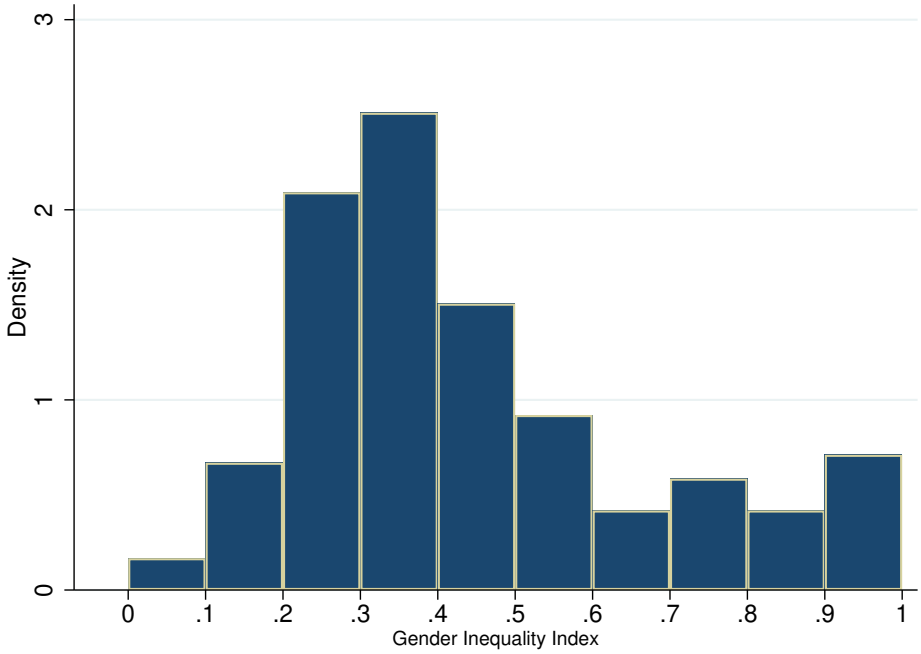
[14]The LEDA algorithm merges ethnicities across dataset through linguistic trees in the Ethnologue. For additional details, see Müller-Crepon *et al.* [Forthcoming]. The algorithm allows us to successfully merge to the Ethnographic Atlas 71% of ethnic groups in the DHS, and 78% of ethnic groups in the Afrobarometer.

Figure 2.2: *Distribution of the eGII in Africa*



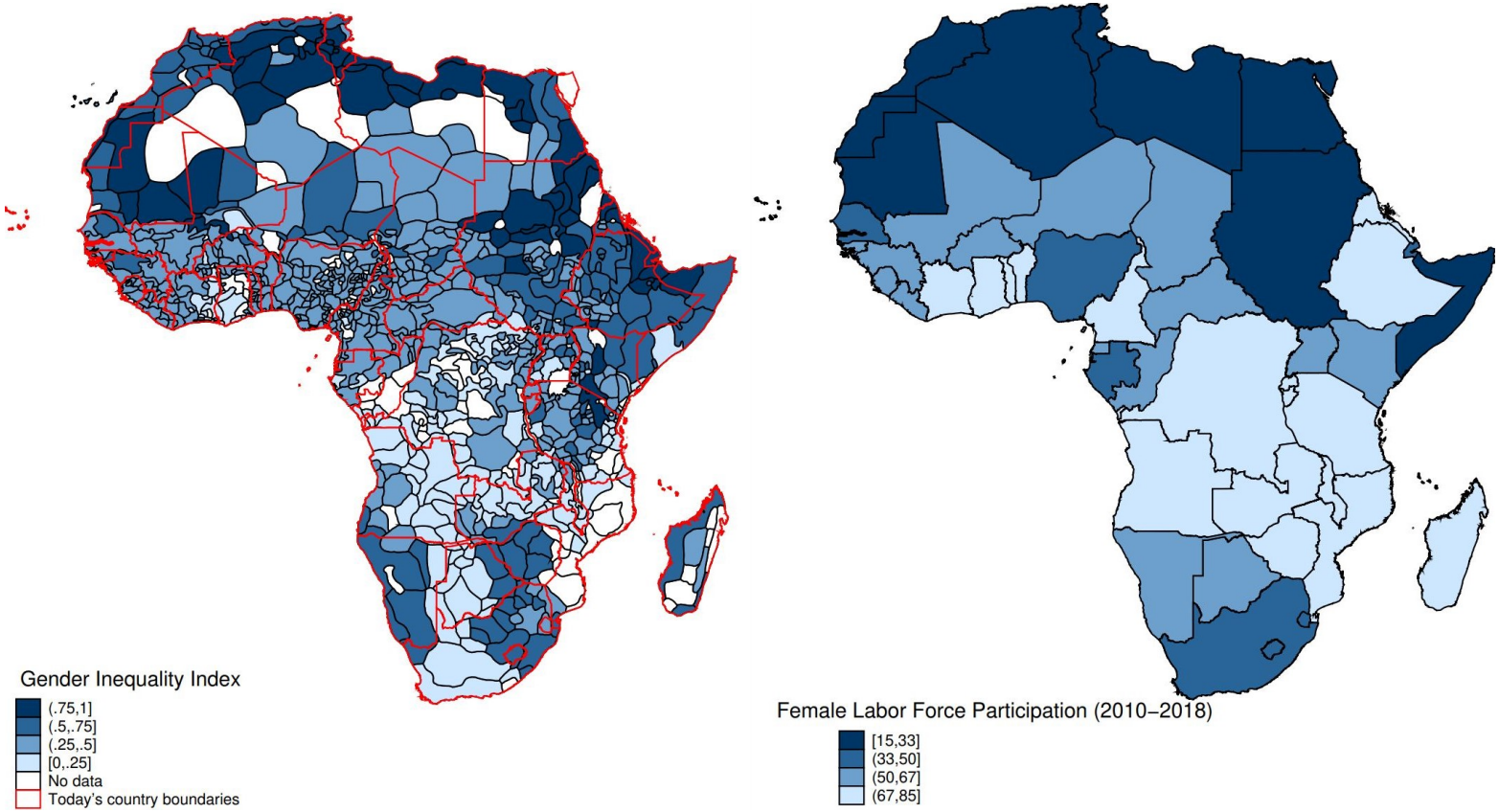
Notes: Mean (standard deviation) of the index: 0.40 (.21).

Figure 2.3: *Distribution of the eGII in our Sample*



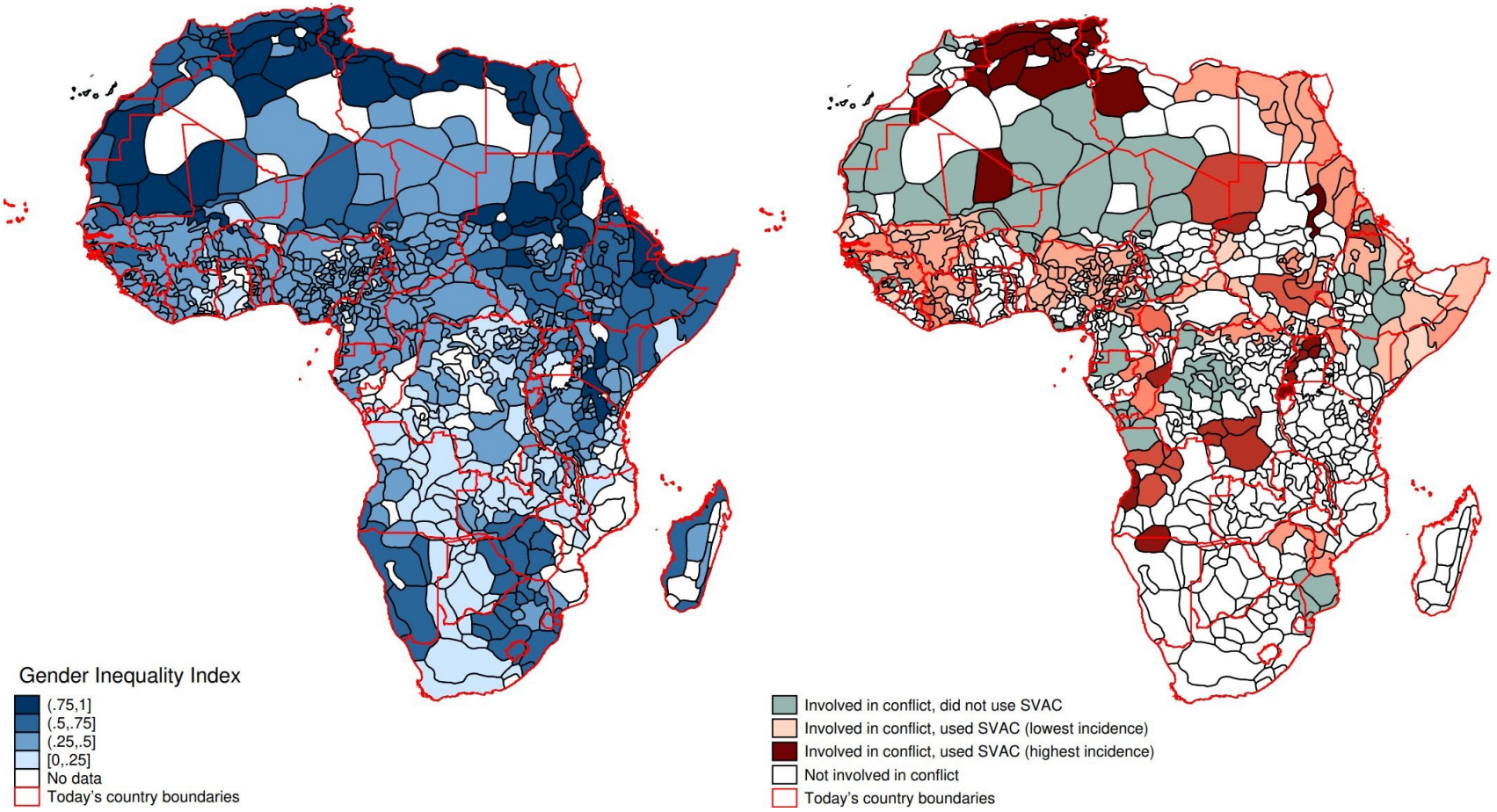
Notes: Mean (standard deviation) of the index: 0.45 (.22).

FIGURE 2.4: *Distribution of the eGII across Africa and Female Labor Force Participation*



NOTES: Left: Distribution of the eGII across Murdock's ethnicities in Africa and contemporary country borders. Right: Female labor force participation at the country level (2010-2018) for women older than 15. Darker colors denote lower participation. Source: International Labor Organization.

Figure 2.5: *Distribution of the eGII across Africa and Sexual Violence in Armed Conflict*



Notes: Left: Distribution of the eGII across Murdock's ethnicities in Africa and contemporary country borders. Right: Total incidence of the use of sexual violence in armed conflict by Murdock ethnicities from 1989 to 2009, measured through an index ranging between 0 and 1.

data is that it allows us to include country fixed effects, which absorb time-invariant country-specific institutions and cultural traits that might have a direct effect on gender norms. We are therefore exploring whether the correlation between our eGII and gender (in)equality today is found across ethnicities within countries.

Tables 2.2 and 2.3 report these correlations. Column (1) in Table 2.2 shows that, within countries in sub-Saharan Africa, women belonging to ethnic groups with high values of the eGII are less likely to work. The eGII is also positively correlated with a measure of son preference (column (2)) and with attitudes justifying intimate partner violence (column (3)). While we do not find a significant correlation between the eGII and severe forms of physical intimate partner violence experienced by women (column (4)), we find a significantly positive correlation between the eGII and sexual violence perpetrated by an intimate partner.

Data from the Afrobarometer survey shows a correlation with some measures of gender attitudes (see Table 2.3). Individuals belonging to ethnic groups with a higher eGII are more likely to agree with the statement that "men make better political leaders than women, and should be elected rather than women" (column (1)) or that "if funds for schooling are limited, a boy should always receive an education in school before a girl" (column (3)). Conversely, respondents belonging to more gender-unequal ethnic groups are less likely to agree with the statement that "women should have the same right as men to own and inherit land" (column (5)). The correlation with other gender attitudes elicited in the survey has the expected sign, but is not statistically significant.[15]

These correlations tend to hold also for the restricted version of the eGII, as shown in Tables B.11 and B.12 in the Appendix. Taken together, this suggests that the eGII—which embeds information on ancestral arrangements that may no longer be in place today—performs fairly well in capturing contemporary measures of gender inequality.

2.5 Gender Inequality and Sexual Violence

Are gender-unequal ethnic groups more likely to be perpetrators of sexual violence in armed conflict when compared to more gender-equal ones? To test our first hypothesis, we estimate the following:

$$SVAC_{ict} = \alpha + \beta eGII_i + \eta_c + \phi_t + \omega_c t + \epsilon_i \tag{2.1}$$

where the dependent variable, $SVAC_{ict}$, denotes the intensity of sexual violence perpetrated by actor i during conflict c in year t. $eGII_i$ captures the gender-inequality index of the perpetrator, either weighted by the size of the EA groups to which each conflict actor corresponds, or unweighted. η_c denotes conflict fixed effects, which account for time-invariant characteristics at the conflict level (e.g., conflict motives, external support, overall conflict cruelty, type of

[15] These are: women and men should have equal rights (column (2)), men should have more right to a job when jobs are scarce (column (4)), it is better for a family if a woman takes care of the home and the children (column (6)).

Table 2.2: *Gender Inequality Index and Gender Inequality Outcomes (DHS)*

	Dependent variable				
				Intimate partner violence:	
	Female employment	Son Preference	Justifies beating	Physical	Sexual
	(1)	(2)	(3)	(4)	(5)
eGII (weighted)	-0.200***	0.063***	0.141**	-0.002	0.078**
	(0.057)	(0.012)	(0.056)	(0.025)	(0.032)
Adj R-squared	0.046	0.157	0.094	0.042	0.041
eGII (unweighted)	-0.227***	0.072***	0.156**	-0.011	0.081**
	(0.063)	(0.013)	(0.061)	(0.029)	(0.038)
Adj R-squared	0.091	0.047	0.157	0.018	0.026
Mean dep. var.	0.580	0.032	0.534	0.064	0.100
Observations	571,184	428,718	481,728	113,192	69,706
Clusters	618	587	564	458	348
Countries	24	24	22	19	15
Years	27	25	20	15	11

Notes: Dependent variables: column (1): female employment; column (2): son preference, defined as (ideal number of boys - ideal number of girls)/(total number of wanted children); column (3) wife beating is justified in at least one of the following instances: she goes out without telling him, she neglects the children, she argues with him, she refuses to have sex with him, she burns the food; column (4) Faced at least one of the following severe physical violence events in the past 12 months: been kicked or dragged; been strangled; been threatened with knife/gun or other weapon; (5) Faced at least one of the following sexual violence events in the past 12 months: physically forced into unwanted sex; forced into other unwanted sexual acts; physically forced to perform sexual acts she didn't want to. Explanatory variables: perpetrator's eGII weighted by the ethnic group land area and unweighted. All explanatory variables are normalized and range between 0 and 1. All columns include country fixed effects and year fixed effects. Standard errors are clustered at the ethnic group's level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table 2.3: *Gender Inequality Index and Gender Attitudes (Afrobarometer)*

	Dependent variable					
	Men better political leaders (1)	Women and men equal rights (2)	Educating boys priority (3)	Men more right to a job (4)	Women right to own land (5)	Women care home and kids (6)
eGII (weighted)	0.109***	-0.057	0.118**	0.022	-0.059**	0.030
	(0.029)	(0.040)	(0.046)	(0.053)	(0.029)	(0.047)
Adj R-squared	0.050	0.062	0.040	0.040	0.117	0.052
eGII (unweighted)	0.144***	-0.074	0.134**	0.045	-0.083**	0.024
	(0.037)	(0.052)	(0.060)	(0.059)	(0.036)	(0.053)
Adj R-squared	0.051	0.062	0.040	0.040	0.117	0.052
Mean dep. var.	0.301	0.717	0.184	0.440	0.740	0.577
Observations	141,567	81,026	36,971	33,420	33,699	32,676
Clusters	770	638	473	413	412	413
Countries	34	34	32	31	31	31
Rounds	5	4	1	1	1	1

Notes: Dependent variables: column (1): agreeing with the statement "Men make better political leaders than women, and should be elected rather than women" as opposed to "Women should have the same chance of being elected to political office as men"; column (2) agreeing with the statement "In our country, women should have equal rights and receive the same treatment as men do" as opposed to "In our country, women should have equal rights and receive the same treatment as men do"; column (3): agreeing with the statement "If funds for schooling are limited, a boy should always receive an education in school before a girl" as opposed to "If funds for schooling are limited, a family should send the child with the greatest ability to learn"; column (4) agreeing with the statement "When jobs are scarce, men should have more right to a job than women"; column (5): agreeing with the statement "Women should have the same rights as men to own and inherit land"; column (6): agreeing with the statement "In general, it is better for a family if a woman has the main responsibility for taking care of the home and children rather than a man". Explanatory variables: perpetrator's eGII weighted by the ethnic group land area and unweighted. All explanatory variables are normalized and range between 0 and 1. All columns include country fixed effects. Columns (1) and (2) include survey round fixed effects. Standard errors are clustered at the ethnic group's level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

warfare, available technology, military tactics). Year fixed effects (ϕ_t) allow to control for time-specific shocks in the whole continent (e.g., the recognition of sexual violence in conflict as a crime, international policies or protocols that might affect data collection and categorization). A conflict-specific year trend ($\omega_c t$) accounts for time-varying factors at the conflict level (e.g., escalation of violence). Standard errors are clustered at the perpetrator level.

The just-described estimating equation abstracts from the victim's characteristics, focusing exclusively on the perpetrator. In order to isolate the role of the perpetrators' eGII from the victim's, we also estimate the following:

$$SVAC_{ijct} = \alpha + \beta eGII_i + \beta eGII_j + \eta_c + \phi_t + \omega_c t + \epsilon_i \tag{2.2}$$

where $SVAC_{ict}$ denotes the intensity of sexual violence perpetrated by actor i against actor j on conflict c and year t, and $eGII_j$ controls for the victim's eGII.

2.5.1 Results

Table 2.4 shows that the eGII is positively and significantly associated with sexual violence. In our preferred specification in column (2), one standard deviation increase in the eGII increases sexual violence by 0.45 standard deviations. When estimating Equation 2.2 and controlling for the victim's eGII, one standard deviation increase in the eGII increases sexual violence by 0.36 standard deviations (column (3)). Columns (4) to (6) show that the same associations hold for the unweighted version of the eGII, and the bottom panel of Table 2.4 shows that coefficients are similar when using the restricted version of the eGII, constructed with only five ethnic characteristics.

Tables B.3 to B.5 in the Appendix repeat this same exercise separately for each of the nine ethnic traits we used to construct the eGII. Interestingly, the association between each characteristic and sexual violence is generally consistent with what discussed in the literature, and with the sign of the loadings in the PCA. Conflict actors with ancestral arrangements conducive to gender equality (matrilineality, stem families, dependence on agriculture) are less likely to perpetrate sexual violence, while actors characterized by more gender-unequal traits (dependence on hunting or gathering, dependence on animal husbandry, or pastoralism) are more likely to perpetrate sexual violence.[16]

2.5.2 Robustness Tests

The just-discussed associations between the eGII and the use of sexual violence in conflict are robust to various alternative specifications. Results are reported in Table B.6 in the Appendix.[17] First, to fully account for the victim's characteristics, we include victim fixed effects in the

[16] In the Appendix, we show that there is an expected association between the slave trade and sexual violence (see section B.2). However, we refrain from including the slave trade in the index due to the fact that it was a geographically constrained historical shock.

[17] Tables B.7 to B.9 are the respective robustness tables for each ethnic characteristic.

Table 2.4: *Gender Inequality Index and Sexual Violence in Armed Conflict*

	Dependent variable: sexual violence (0-3)					
	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic Gender Inequality Index						
eGII (weighted)	1.83***	1.87***	1.55**			
	(0.505)	(0.564)	(0.715)			
eGII (unweighted)				1.84***	1.90***	1.54**
				(0.457)	(0.507)	(0.635)
Adjusted R^2	0.307	0.377	0.376	0.311	0.382	0.382
restricted eGII (weighted)	1.40***	1.43***	1.45*			
	(0.502)	(0.525)	(0.755)			
restricted eGII (unweighted)				1.54***	1.57***	1.46**
				(0.455)	(0.480)	(0.658)
Adjusted R^2	0.299	0.368	0.369	0.307	0.376	0.377
Mean dep. var.	0.62	0.62	0.62	0.62	0.62	0.62
Conflict fixed effect	yes	yes	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes	yes	yes
Conflict-specific time trend		yes	yes		yes	yes
Victim's eGII			yes			yes
Observations	900	900	893	900	900	893
Clusters	128	128	127	128	128	127

Notes: OLS coefficient estimates of Equation 2.1. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. The explanatory variables are the perpetrator's eGII (weighted by the ethnic group land area and unweighted) and the perpetrator's restricted version of the eGII (weighted by the ethnic group land area and unweighted). All explanatory variables are normalized and range between 0 and 1. Standard errors are clustered at the perpetrator's level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

main specification. Columns (1) and (2) in Tables B.7 to B.9 display coefficients of this—more demanding—specification, for the weighted and unweighted version of the eGII, respectively. The significance of the estimates tends to fall, although the coefficients maintain the expected sign.

Next, we run Equation 2.1 replacing conflict fixed effects with country fixed effects. This specification is less conservative than our preferred one, since one country may experience multiple conflicts.[18] However, in few instances, one conflict may span across multiple countries.[19]

[18] Between 1989 and 2009, for example, Niger experienced what UCDP-GED defines as three different conflicts, i.e., the first, second, and third Tuareg rebellions, respectively, fought by five different rebel groups against the government.

[19] For example, the conflict between the government of the Central African Republic and the Forces of Francoise Bozize took place both in the Central African Republic and in Chad. Similarly, some events in the conflict between the government of Ethiopia and the Oromo Liberation Front took place in Kenya.

As columns (3) and (4) show, the coefficients are insensitive to the choice of fixed effects, and maintain the same magnitude and significance as in the main specification.

Finally, in columns (5) and (6) we abstract from the temporal variation present in our data. Since the independent variables—i.e., ethnic ancestral characteristics—are time-invariant, we collapse the data and have as unit of observation a dyad (perpetrator-victim pair) in a specific conflict and country. The outcome variable is the average sexual violence intensity perpetrated by each actor in all years of a specific conflict. Reassuringly, results are similar to those obtained with the specification that includes the temporal variation.

2.6 Cultural Distance in Gender Norms and Sexual Violence

Does cultural distance in gender norms between the combatants help explain the emergence of sexual violence in ethnic conflicts? To test our second hypothesis, we take advantage of the dyadic structure of the data.[20]

For each actor in every year of conflict, we have information on their ethnic characteristics, their use of sexual violence and, most importantly, their opponents. Two actors fighting against each other in a conflict constitute a *dyad*. For instance, the government of Chad (corresponding to the "Zaghawa, Bideyat" ethnic group) and the *Comité de Sursaut National pour la Paix et la Démocratie* (CSNPD) rebel group (corresponding to the "Sara" ethnic group) form a dyad in our dataset. Since we have information on sexual violence perpetrated by both actors, our dyadic dataset is bidirectional. Therefore, every dyad involved in a conflict event appears twice in the dataset. In one instance, the government of Chad is the perpetrator and the CSNPD is the victim. The variable $SVAC$, in this case, captures the intensity of sexual violence inflicted by the government of Chad to the CSNPD. In a second instance, the government of Chad is the victim, and the CSNPD is the perpetrator. In this case, the variable $SVAC$ captures the intensity of sexual violence inflicted by the rebel group CSNPD to the government of Chad.

We first construct a measure of absolute distance between the perpetrator's gender inequality index and the victim's gender inequality index as follows:

$$\text{eGII}_{pv}^{Dist} = |\text{eGII}_p - \text{eGII}_v|$$

Next, we estimate the following specification, in the spirit of a gravity approach (see Grosjean [2011] and Serafinelli and Tabellini [2017] as examples of gravity equations applied to culture):

$$\text{SVAC}_{pvct} = \alpha + \gamma \text{eGII}_{pv}^{Dist} + \Phi_c + \tau_t + \omega_c t + P_p + \epsilon_{pvct} \tag{2.3}$$

The dependent variable is an index capturing the intensity of sexual violence perpetrated by actor p against actor v during conflict c in year t; Φ_c and τ_t denote conflict and year fixed effect, respectively, and $\omega_c t$ is a conflict-specific time trend. The inclusion of perpetrator fixed effect (P_p) allows to control for perpetrator-specific time invariant characteristics such as own gender

[20] Table B.2 reports an extract of our dataset.

inequality index, overall aggressiveness, and other ethnic traits. Standard errors are clustered at the dyadic level. We restrict the sample to inter-ethnic conflicts only, i.e., to those where we are able to assign distinct ethnic identities to the perpetrator and the victim.

To further explore the nature of the cultural clash, we split the absolute distance measure into two components, and separately assess the impact of (i) the perpetrator being *more gender unequal* than the victim and (ii) the perpetrator being *less gender unequal* than the victim when explaining the use of sexual violence in a conflict:

$$\text{Perpetrator More Unequal}_{pv} = \begin{cases} |\text{eGII}_p - \text{eGII}_v| & \text{if eGII}_p > \text{eGII}_v \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{Perpetrator Less Unequal}_{pv} = \begin{cases} |\text{eGII}_p - \text{eGII}_v| & \text{if eGII}_p < \text{eGII}_v \\ 0 & \text{otherwise.} \end{cases}$$

To tease out the separate effect of these two distinct components of cultural distance, we estimate the following:

$$\begin{aligned} \text{SVAC}_{pvct} = \alpha + \eta_1 &\text{Perpetrator More Unequal}_{pv} \\ &+ \eta_2 \text{Perpetrator Less Unequal}_{pv} + \Phi_c + \tau_t + \omega_c t + P_p + \epsilon_{pvct} \end{aligned} \quad (2.4)$$

This specification is equivalent to the one in Equation 2.3, and it differs only in that it substitutes the cultural distance measure with its two main components. In this specification, η_1 and η_2 separately capture the effects of two distinct cultural clashes: one where the perpetrator faces an opponent characterized by more gender-equal cultural norms compared to its own norms (η_1), and one where the perpetrator is confronted with an opponent characterized by more gender-unequal cultural norms (η_2).

2.6.1 Results

Results are reported in Table 2.5. Column 1 presents the coefficient estimate of γ in Equation 2.3. There is a positive and significant association between the absolute cultural distance in gender norms of the combatants and the intensity in sexual violence in conflict. One standard deviation increase in the absolute distance in the eGII of the combatants increases the intensity of sexual violence by 0.21 standard deviations.

Columns (2)-(5) unpack this association, and separately assess the role played by the perpetrator's own eGII and by two distinct cultural clashes: when (i) the perpetrator is confronted with an opponent who holds more gender-equal cultural norms (*Perpetrator more unequal*) (ii) the perpetrator is confronted with an opponent that holds more gender-unequal norms (*Perpetrator less unequal*). Column (2) displays results of a horse-race between the perpetrator's eGII and the absolute cultural distance when the perpetrator is more gender unequal than the victim. The coefficient on eGII is positive, but statistically insignificant and smaller than in Table 2.4, which showed the association between the eGII alone and sexual violence. Instead,

Table 2.5: *Cultural Distance in Gender Norms and Sexual Violence in Armed Conflict*

	Dependent variable: sexual violence (0-3)				
	(1)	(2)	(3)	(4)	(5)
Absolute distance ($\lvert eGII_p - eGII_v \rvert$)	1.53***				
	(0.518)				
Perpetrator's eGII		0.58	2.05**	1.13	
		(0.629)	(0.957)	(0.997)	
Perpetrator *more* unequal		1.44*		1.51*	1.53***
		(0.814)		(0.811)	(0.503)
Perpetrator *less* unequal			0.20	0.64	1.56
			(0.919)	(0.888)	(1.214)
Conflict fixed effect	yes	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes	yes
Conflict-specific time trends	yes	yes	yes	yes	yes
Perpetrator fixed effect	yes				yes
Mean dep. var.	0.62	0.62	0.62	0.62	0.62
Observations	623	643	643	643	623
Clusters	76	76	76	76	76
Adjusted R^2	0.597	0.379	0.374	0.379	0.596

Notes: OLS coefficient estimates of Equations 2.3 and 2.4. The sample is restricted to include inter-ethnic conflicts only. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. The explanatory variables are the following: the absolute distance in the eGII between perpetrator and victim; the perpetrator's eGII (weighted by the ethnic group land area); the absolute distance in the eGII between perpetrator and victim when the perpetrator is more gender unequal than the victim; the absolute distance in the eGII between perpetrator and victim when the perpetrator is less gender unequal than the victim. Standard errors are clustered at the dyad level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

the coefficient on *Perpetrator more unequal* is larger in magnitude, and significant at the 10 percent level. Column 3 includes *Perpetrator less unequal*, i.e., the absolute cultural distance when the perpetrator holds more equitable gender norms than the victim. The latter factor seems to not be positively associated with the use of sexual violence: the coefficient is small and not significantly different from zero. Instead, the coefficient on the perpetrator's eGII is large and significant in this specification.

Column (4) shows coefficient estimates for a specification that simultaneously includes the perpetrator's eGII and the two different cultural clashes. All coefficients estimates are positive, but the largest and only significant one is that on the perpetrator being more gender-unequal than the victim. This coefficient can be interpreted as follows: when a gender-unequal perpetrator with a eGII of 1 faces a gender-egalitarian victim with a eGII of 0, sexual violence intensity is 1.51 higher than when the perpetrator and the victim hold the same gender norms. In column (5), instead of controlling for the perpetrator's eGII, we add perpetrator fixed effects. Crucially, this allows us to account for any time-invariant perpetrator's characteristics that may confound the results, like other ethnic traits (including the perpetrator's own eGII),

overall aggressiveness, whether the perpetrator is a state force or a rebel group. The coefficient barely changes in magnitude, and its statistical significance increases.

Interpretation of the Results

So far, we have interpreted the positive and significant coefficient of *Perpetrator more unequal* in Table 2.5 as the result of a cultural clash between combatants. However, one potential alternative explanation is that perpetrators strategically use sexual violence to target a valuable asset (i.e., women) in the opponent's society. This behavior would arise when women have a prominent role in victim's society—and consequently, it would be more likely when perpetrators are more gender-unequal than the victim. This could occur also in the absence of a cultural clash driven by divergent gender norms.

To rule out this alternative explanation, we re-run our main specification controlling for victim's characteristics, which take into account the economic value of women in victim's society—as well as any other time-invariant victim's characteristics.[21] By and large, our results are robust to this procedure (see Table B.13). When controlling for the victim's eGII rather than the perpetrator's, the coefficient on *Perpetrator more unequal* stays large and significant. When adding victim's fixed effects, the same coefficient loses significance despite maintaining a similar magnitude.

Another threat to the interpretation of the results could be sexual violence misreporting. The positive coefficient on *Perpetrator more unequal* could reflect the fact that victims from more gender-equal societies may be likely to report incidents of sexual violence. Again, the robustness of our results to the inclusion of victim fixed effects indicates that misreporting is not driving our results.

These tests support our hypothesis that cultural distance is what drives sexual violence in conflict, and not combatant-specific factors. A subsequent natural question to ask would be: what is the mechanism by which cultural distance triggers sexual violence? We propose a new explanation based on the identity model by Akerlof and Kranton [2000]. Perpetrators that confront more gender-equal opponents might perceive that the relatively better position of women is a threat to their own norms or ideals, and might resort to sexual violence to reinstate the loss of utility suffered by this encounter. The suggested explanation is consistent with the asymmetry of our findings, that is, the fact that we only find an association when the perpetrator is more gender-unequal than the victim, but not viceversa. Perpetrators that face a more gender-unequal opponent might not feel that the worse relative position of women in the opponents' societies threatens their masculinity. Since their identity utility is not affected, they do not necessarily respond with sexual violence to the clash in cultural norms.

To further validate our results, in the next Subsections we conduct a battery of robustness checks, and explore the relationship between gender-norms distance and general (i.e.,

[21] For instance, whether women on the victim's side are a particularly easy target. It is not clear how this would be related to gender inequality: on the one hand, women in gender-equal societies might move more freely in public spaces but, on the other hand, women in gender-unequal societies might be more vulnerable.

non-sexual) violence as well as the relationship between other measures of cultural distance—linguistic and religious—and sexual violence in conflict.

2.6.2 Robustness Checks

We report robustness checks in Tables B.14 to B.17 in the Appendix. Given that our measure of cultural distance in gender norms is time-invariant, we show that our results are similar when running the same specifications abstracting from the temporal variation in the data (Table B.14).

Table B.15 shows that the coefficients are robust to the inclusion of alternative sets of fixed effects, and to alternative versions of the eGII. In column (1), we include conflict-year fixed effects, to account for any conflict-year specific factor that may confound the results (e.g., how cruel the conflict was in that specific year, changes in military tactics during the conflict, and so on). In column (2), we add country fixed effects. Our results remain unchanged when using the unweighted version of the eGII, and when using the restricted version of the eGII that includes only five ethnic characteristics.

We also test whether our results hold when assigning an alternative measure of the eGII to state military forces. This exercise is motivated by the fact that the composition of the state military may not reflect the ethnic identity of the government. For instance, it is possible that the election of a new government does not (at least immediately) result into an alignment of the military with the ethnic identity of the new ethnic groups in power. For this reason, we explore whether our results hold when assigning to the government a more conservative measure, i.e. the average eGII of all the Murdock ethnic groups within a country, weighted by the size of each group's settlement area. Table B.16 shows that replacing the government's eGII with the country average leaves the results almost unchanged: the *Perpetrator more unequal* coefficient stays significant and slightly increases in magnitude compared to our baseline specifications in Table 2.5. Conversely, the *Perpetrator less unequal* coefficient remains insignificant, and its size slightly decreases.

Finally, in Table B.17, we show that results are robust to multi-way clustering, i.e. to clustering standard errors at the level of the first and of the second actor in a given pair. This allows for arbitrary correlations of the error term within a group of actors pairs that share the same perpetrator or that share the same victim.

2.6.3 Gender-Norms Clash and General Violence

Cultural distance in gender norms may be correlated to other dissimilarities between ethnic groups, which could in turn generate more violent conflicts, or more violent episodes within a conflict. If this was the case, sexual violence would only be a byproduct of general violence, and the cultural clash we are estimating would not be specifically linked to a gender-based form of violence.

The inclusion of conflict fixed effects in our main specification partially alleviates this concern because it accounts for the overall cruelty in a conflict. However, to fully rule out this alternative explanation, we run a placebo test using a different measure of violence as the outcome variable. We exploit information on the number of fatalities experienced by a conflict actor in every year of conflict to construct an index similar to the sexual violence variable, ranging between 0 to 3 depending on the number of recorded deaths. Since this measure of general violence is also bidirectional, we can run the same specifications of Equations 2 and 3 having as an outcome variable the number of deaths inflicted by the perpetrator on the victims.

Table 2.6: *Distance in Gender Norms and General Violence*

	Dependent variable: victim's fatalities (0-3)				
	(1)	(2)	(3)	(4)	(5)
Absolute distance ($\lvert eGII_p - eGII_v \rvert$)	-0.18				
	(0.856)				
Perpetrator's eGII		-0.03	-1.40	-0.63	
		(0.720)	(0.991)	(1.136)	
Perpetrator *more* unequal		-1.17		-1.24	-0.06
		(0.807)		(0.805)	(0.976)
Perpetrator *less* unequal			-0.35	-0.71	-0.58
			(1.011)	(0.982)	(1.111)
Conflict fixed effect	yes	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes	yes
Conflict-specific time trends	yes	yes	yes	yes	yes
Perpetrator fixed effect	yes				yes
Mean dep. var.	1.07	1.07	1.07	1.07	1.07
Observations	623	643	643	643	623
Clusters	76	76	76	76	76
Adjusted R^2	0.317	0.266	0.263	0.265	0.316

Notes: OLS coefficient estimates of Equations 2.3 and 2.4. The sample is restricted to include inter-ethnic conflicts only. The dependent variable is an index ranging between 0 and 3 that captures the intensity of deaths inflicted by the perpetrator on the victim, coded like the sexual violence variable (0: no killings; 1: between 1 and 24; 2 between 25 and 999; 3: equal to or larger than 1000). The explanatory variables are the following: the absolute distance in the eGII between perpetrator and victim; the perpetrator's eGII; the absolute distance in the eGII between perpetrator and victim when the perpetrator is more gender unequal than the victim; the absolute distance in the eGII between perpetrator and victim when the perpetrator is less gender unequal than the victim. Standard errors are clustered at the dyad level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table 2.6 reports the results of this exercise. Column (1) shows that the distance in gender norms of the combatants is not associated with the number of deaths inflicted to the victims: the coefficient is small and not statistically different from zero. The same holds for the coefficients in columns (2) to (5). Neither the perpetrator's own gender norms, nor the cultural clashes are positively and significantly associated with general violence. If at all, these elements seem to be

negatively associated with general violence, although none of these coefficients is statistically significant. Taken together, this suggests that the perpetrator's eGII and the cultural clash in gender norms between perpetrator and victim explain only the use of *gender-based* forms of violence.

2.6.4 Distance in Gender Norms and Other Measures of Cultural Distance

In this section, we assess whether combatants' clashes in gender norms are the main driver of the use of sexual violence, or whether clashes in other cultural dimensions are similarly powerful in explaining this phenomenon. To disentangle the role of gender norms from other aspects of culture, we exploit linguistic and religious distance, widely-used proxies for cultural differences between populations.[22]

We use Fearon's [2003] measure of linguistic distance (called cladistic distance), which is based on linguistic trees in the Ethnologue, a comprehensive database of more than 7,000 known living languages. We merge information on languages spoken by ethnic groups through the Ethnic Power Relations-Ethnic Dimensions (EPR-ED) dataset, and compute distances between each pair of languages based on the number of common nodes in the tree. This allows us to compute a measure of linguistic distance between ethnic groups, and ultimately, between the perpetrator and the victim. Following the same methodology, we construct a measure of religious distance. Sections B.1.5 and B.1.6 in the Appendix provide additional details on how we construct these measures.

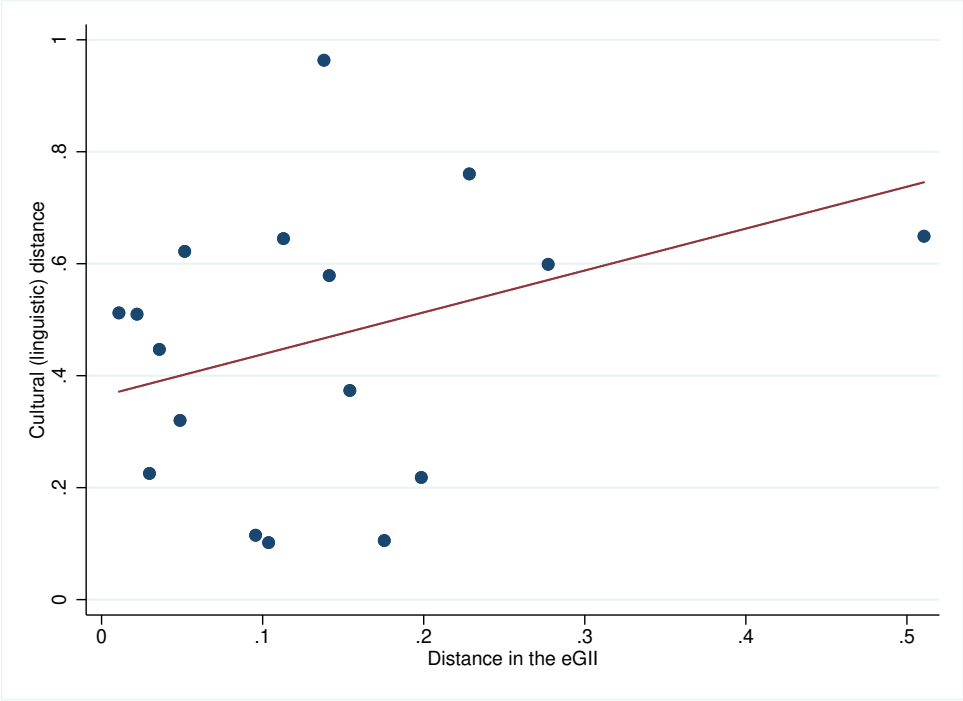
Figure 2.6 plots the correlation between linguistic distance between the combatants and our measure of distance in gender norms. Not surprisingly, the correlation is positive, suggesting that conflict actors that are linguistically distant are on average also more likely to differ in their gender norms. However, the figure and the respective correlation coefficient (0.25, statistically significant at the one percent level) also suggest that distance in gender norms is not a perfect predictor of overall cultural distance. On the other hand, distance in gender norms is uncorrelated with religious distance (see Figure 2.7).[23] Taken together, this suggests that our gender norms measure captures a dimension distinct from already-proposed aspects of cultural distance.

To assess whether cultural clashes in traits unrelated to gender norms are associated with sexual violence, we first regress linguistic distance on distance in gender norms, and obtain residuals of this regression. These residuals capture the component of cultural distance that is unexplained by differences in gender norms. We re-run our main specifications controlling for this component of cultural distance and, alternatively, for linguistic distance. As can be seen in columns (2)-(3) and columns (5)-(6), this leaves the coefficients in our main specifications

[22]Spolaore and Wacziarg [2016a] use other country-level measures of cultural heterogeneity, like answers to World Value Survey (WVS) questions. We abstain from using the WVS due to potential reverse causality between this measure and our outcome variable. In addition, Spolaore and Wacziarg [2016a] proposed genetic distance as a summary measure for populations' *relatedness*. This measure is unfeasible in our context due to a too wide categorization of ethnic groups in the original source of genetic data.

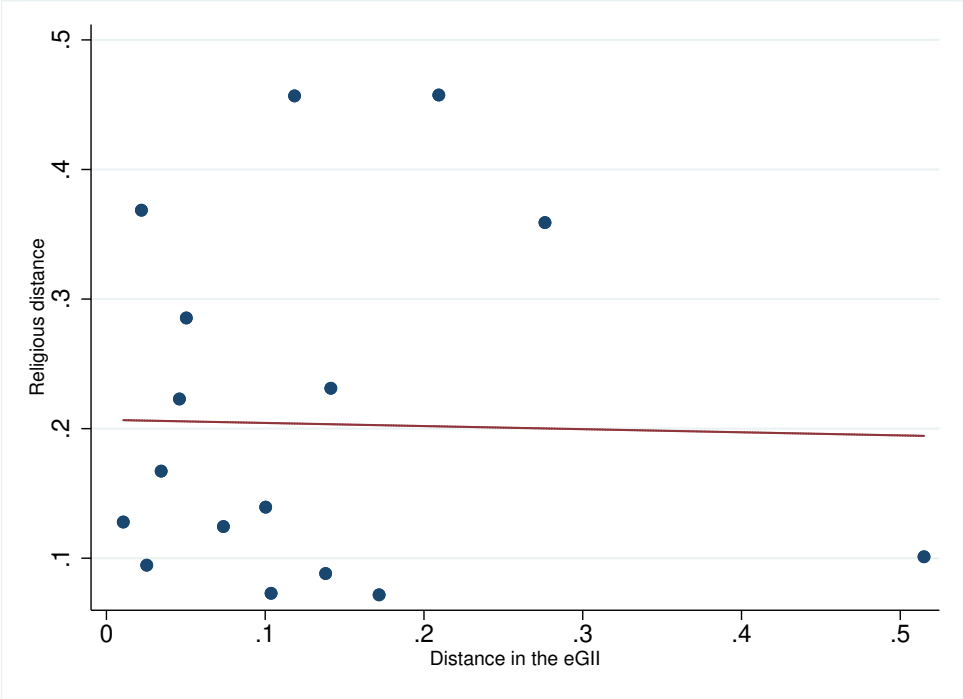
[23]Religious distance and linguistic distance are positively correlated. See Figure B.6 in the Appendix.

Figure 2.6: *Correlation between Cultural Distance in Gender Norms and Linguistic Distance*



Notes: Correlation between the absolute distance in gender norms between the combatants and their cultural (linguistic) distance for the sample of ethnicities involved in inter-ethnic conflict. Correlation coefficient: 0.25***. Sources: Murdock Ethnographic Atlas and Ethnologue.

Figure 2.7: *Correlation between Cultural Distance in Gender Norms and Religious Distance*



Notes: Correlation between the absolute distance in gender norms between the combatants and their religious distance for the sample of ethnicities involved in inter-ethnic conflict. Correlation coefficient: -0.02. Sources: Murdock Ethnographic Atlas and EPR-ED dataset.

Table 2.7: *Distance in Gender Norms, Linguistic Distance, and Sexual Violence in Conflict*

	Dependent variable: sexual violence (0-3)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance in gender norms ($\lvert eGII_p - eGII_v \rvert$)	1.53***	1.52***	1.54***					
	(0.518)	(0.513)	(0.516)					
Perpetrator *more* gender unequal				1.53***	1.51***	1.53***		
				(0.504)	(0.502)	(0.500)		
Perpetrator *less* gender unequal				1.56	1.54	1.56		
				(1.214)	(1.205)	(1.209)		
Distance in other cultural traits (residuals)		-0.02			-0.02		-0.08	
		(0.050)			(0.049)		(0.107)	
Linguistic distance			-0.02			-0.02		-0.01
			(0.050)			(0.049)		(0.072)
Conflict fixed effect	yes	yes	yes	yes	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes	yes	yes	yes	yes
Conflict-specific time trends	yes	yes	yes	yes	yes	yes	yes	yes
Perpetrator fixed effect	yes	yes	yes	yes	yes	yes	yes	yes
Mean dep. var.	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62
Observations	623	623	623	623	623	623	623	623
Clusters	76	76	76	76	76	76	76	76
Adjusted R^2	0.597	0.596	0.596	0.596	0.596	0.596	0.592	0.592

Notes: OLS coefficient estimates of Equations 2.3 and 2.4 with controls for distance in other cultural traits or linguistic distance. The sample is restricted to include inter-ethnic conflicts only. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. The explanatory variables are the following: the absolute distance in the eGII between perpetrator and victim; the absolute distance in the eGII between perpetrator and victim when the perpetrator is more gender unequal than the victim; the absolute distance in the eGII between perpetrator and victim when the perpetrator is less gender unequal than the victim; residuals of regressing linguistic distance on distance in gender norms; linguistic distance between perpetrator and victim. Standard errors are clustered at the dyad level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table 2.8: *Distance in Gender Norms, Religious Distance, and Sexual Violence in Conflict*

	Dependent variable: sexual violence (0-3)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance in gender norms ($\lvert eGII_p - eGII_v \rvert$)	1.53***	1.65***	1.65***					
	(0.518)	(0.552)	(0.556)					
Perpetrator *more* gender unequal				1.53***	1.96***	1.96***		
				(0.504)	(0.465)	(0.470)		
Perpetrator *less* gender unequal				1.56	1.14	1.14		
				(1.214)	(1.114)	(1.114)		
Distance in other cultural traits (residuals)		-0.06			0.03		-0.44**	
		(0.182)			(0.177)		(0.212)	
Religious distance			-0.06			0.03		-0.45**
			(0.182)			(0.177)		(0.209)
Conflict fixed effect	yes	yes	yes	yes	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes	yes	yes	yes	yes
Conflict-specific time trends	yes	yes	yes	yes	yes	yes	yes	yes
Perpetrator fixed effect	yes	yes	yes	yes	yes	yes	yes	yes
Mean dep. var.	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62
Observations	623	590	590	623	590	590	590	590
Clusters	76	72	72	76	72	72	72	72
Adjusted R^2	0.597	0.560	0.560	0.596	0.559	0.559	0.556	0.556

Notes: OLS coefficient estimates of Equations 2.3 and 2.4 with controls for distance in other cultural traits or religious distance. The sample is restricted to include inter-ethnic conflicts only. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. The explanatory variables are the following: the absolute distance in the eGII between perpetrator and victim; the absolute distance in the eGII between perpetrator and victim when the perpetrator is more gender unequal than the victim; the absolute distance in the eGII between perpetrator and victim when the perpetrator is less gender unequal than the victim; residuals of regressing religious distance on distance in gender norms; religious distance between perpetrator and victim. Standard errors are clustered at the dyad level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

almost unchanged. Columns (7) and (8) in Table 2.7 show that neither overall cultural distance nor residuals alone can explain actors' use of sexual violence in conflict (the coefficient is small in magnitude and insignificant).

Table 2.8 repeats the same exercise using religious distance as a proxy for cultural distance between perpetrator and victim. Again, controlling for residuals or religious distance does not affect our main results. Interestingly, column (8) shows that, if at all, religious distance is *negatively* associated with the use of sexual violence.[24] Taken together, these results suggest that what matters in explaining the use of sexual violence is not cultural distance in general, but a specific clash in cultural norms related to gender.

2.7 Conclusion

Why do some conflict actors systematically rape, while others never do so? In this paper, we advance and test a new hypothesis for the use of sexual violence in armed conflict.

We find that armed actors characterized by more gender-unequal norms are more likely to engage in sexual violence during ethnic conflict. However, we show that this explanation for sexual violence—stemming from gender inequality on the perpetrator's side—is incomplete. The prevalence and the intensity of war-related sexual violence is better explained when considering both the perpetrator's and the victim's gender norms. In particular, sexual violence emerges and intensifies when there is a *clash of conceptions* between combatants on what is the appropriate role of men and women in society. Cultural distance in gender norms between perpetrator and victim explains sexual violence more strongly than the perpetrator's own gender inequality.

When examining this relationship further, we uncover that the effect is driven by a specific cultural clash, i.e., when the perpetrator holds more gender-unequal norms than the victim. We show that this just-described pattern is specific to gender-based violence, and that it does not explain the intensity of general violence perpetrated by an armed actor, measured by deaths inflicted on the opponent. Moreover, conflict-related sexual violence is not driven by general cultural differences, but by differences in gender norms.

Our contribution in this paper is threefold. First, we enrich existing conflict data sources with ethnic characteristics of the groups involved. This novel dyadic dataset may constitute a potentially valuable resource for future contributions in the conflict literature. While the literature has mainly focused on understanding how ethnic differences determine the onset of conflicts, we show that they can also explain how violence manifests once the conflict has started. Among the different ethnic dimensions that can induce violence, we focus on the role of gender norms in explaining the occurrence and intensity of sexual violence in conflict. Future research should continue to uncover how different ethnic dimensions can trigger and exacerbate conflict.

Second, we propose and validate an ethnic gender inequality index at the ethnic-group level for Africa. This index complements the gender inequality index (GII) introduced in 2010 by the

[24] Given that distance in gender norms and religious distance are uncorrelated, it is not surprising that the coefficient in column 7 is almost identical to the one in column 8.

United Nations Development Programme, which is constructed at the country level and based on contemporary variables (reproductive health, empowerment and labor market participation). Our Gender Inequality Index is instead constructed at the ethnic group level, is based on anthropological notions of gender (in)equality, and aims at capturing the deeply entrenched norms of a society. The large within-country variation found with our eGII makes more valuable our exercise, especially for regions in the world, such as the African continent, with high ethnic diversity.

Finally, in line with recent literature on the cultural determinants of intimate-partner violence (Alesina *et al.* [2021]; Tur-Prats [2019]), we find that violence against women during wartime shares the same fundamental causes as violence against women during peacetime. From a policy perspective, this suggests that policies aimed at changing gender norms might have an effect on all the different manifestations of violence against women.

Chapter 3

Cultural Distance and Ethnic Civil Conflict

3.1 Introduction

Since the end of World War II, civil conflicts have been the most common form of war around the world. Most civil conflicts have taken place in Africa, and the deadliest have been fought along ethnic lines.[1] During one year of the Rwandan Civil War between the Hutu and the Tutsi, at least 500,000 people were killed (Sundberg and Melander [2013]). Globally, civil conflict has caused five times as many deaths as inter-state war, and has led to long-lasting economic and social disruptions (Fearon and Laitin [2003]). Yet, what triggers ethnic civil conflict remains largely debated. A well-established literature has analyzed the role of ethnic-diversity measures at the country level using ethnic fractionalization (Fearon and Laitin [2003], Esteban *et al.* [2012]) and polarization indices (Montalvo and Reynal-Querol [2005], Esteban *et al.* [2012]), and has found that more diverse countries tend to experience more conflict. However, given that all ethnicities in a country face the same aggregate level of diversity, why do some rebel against the central government while others do not?

To answer this question, I proceed in two steps. First, I move the study of diversity and conflict from the country level to the ethnicity level and explore how ethnic groups' cultural *distance* to the central government affects their decision to rebel against it. Second, in line with existing theories of conflict (Esteban and Ray [2011], Spolaore and Wacziarg [2017]), I distinguish between conflicts fought over government power (i.e., public goods) and conflicts fought over territory and resources (i.e., rival goods).[2] I test the hypothesis that cultural distance—

[1] Ethnic civil conflicts, the focus of this paper, are "armed conflicts between the government of a state and one or more internal opposition groups that cause at least 25 battle-related deaths within a year in which armed groups (i) explicitly pursue ethno-nationalistic aims, motivations, and interests; and (ii) recruit fighters and forge alliances on the basis of ethnic affiliation" (Wimmer *et al.* [2009]).

[2] The data source I use in this paper categorizes each event into one of these two conflict types, i.e., either over government power or territory. Between 1961 and 2017, 34 percent of the whole prevalence of ethnic conflict in Africa was accounted for by conflict over territory, and 66 percent by conflict over power (Sundberg and Melander [2013]).

defined as differences in beliefs, values, and preferences (Mokyr [2016])—should only explain conflicts fought over government power. By nature, every group in a country is subjected to public policies provided by the government. Different preferences over public policies between an ethnicity and the government may give rise to disagreement and ultimately conflict. Therefore, a large cultural distance to the government should increase an ethnicity's propensity to fight over government power, but not its propensity to fight over territory. If anything, ethnicities with different preferences and tastes should be less likely to compete over the same type of rival good.

While prominent in the theoretical literature on diversity and conflict (Esteban and Ray [2011], Caselli and Coleman [2013], Spolaore and Wacziarg [2017]), this hypothesis has never been empirically tested at the ethnic-group level in the context of civil war. To do so, I generate a novel dataset measuring how culturally distant each ethnic group is from the central government at a given point in time. I measure cultural distance using linguistic distance, as is standard in the literature.[3] I follow all ethnic groups over time and construct an indicator for whether they rebel against the government in a specific year or not. I then classify each rebellion as either *conflict over government power* or *conflict over territory*. The resulting dataset is a panel of 236 distinct ethnic groups in 43 African countries observed over a period of 57 years (1961-2017).[4]

In the empirical analysis, I leverage changes in the ethnic identity of the government as a source of within-group time variation in cultural distance, and find support for my hypothesis. After a leadership change, the prevalence of conflict over power increases among ethnic groups that become culturally more distant to the government, and decreases among ethnic groups that become culturally closer. These effects cannot be attributed to differential trends in conflict prior to a government change across these two subgroups of ethnicities. The reaction to cultural distance occurs immediately after a change in leadership, and is large in magnitude: a one standard deviation increase in cultural distance increases the prevalence of conflict over government power by 0.36 standard deviations.[5] Conversely, I do not find a significant association between cultural distance to the government and conflict over territory.

My results hold after conditioning on a rich set of fixed effects. The data allows me to include ethnicity fixed effects, thus isolating the effect of cultural distance from any time-invariant ethnic-specific characteristic potentially associated with conflict: cultural norms, social structure, ancestral arrangements and geographic conditions, whether the group was split by country borders, or a general propensity to rebel. Since the source of within-ethnicity variation

[3]See, for example, Fearon and Laitin [2003], Desmet *et al.* [2009], Desmet *et al.* [2012], Esteban *et al.* [2012]. The use of linguistic distance is motivated by the notion that language is a salient dimension of culture transmitted through generations (Spolaore and Wacziarg [2016a]). Different languages are the result of horizontal separations between populations, and these separations are likely to go hand in hand with cultural divergence.

[4]I focus on Africa for two reasons. First, it constitutes a unique setting due to the high degree of ethnic diversity within countries, which leads to the highest-quality data on ethnicity (e.g., the Linking Ethnic Data for Africa (LEDA) allows me to link several datasets on ethnicity and is only available for Africa). Second, for the purpose of identification, Africa offers two unique natural experiments I exploit in my analysis: the Bantu expansion and the random allocation of country borders.

[5]For comparison, this corresponds to three times the effect of an increase in resource inequality uncovered by Guariso and Rogall [2017]. They show that a one standard deviation increase in rainfall inequality between an ethnicity and the government increases the likelihood of conflict by 0.12 standard deviations.

in cultural distance results from government changes, the inclusion of country-by-year fixed effects ensures that the coefficients are not confounded by government characteristics, or by country-level time shocks affecting all ethnicities within a country, including the overall effect of a leadership change. Moreover, country-by-year fixed effects allow me to keep constant the aggregate level of diversity (e.g., a country's level of ethnic fractionalization or polarization) and focus on the effect of cultural distance between groups.

While I interpret my findings as the effect of cultural distance on conflict, I consider alternative explanations. First, I show that the effect of cultural distance is not confounded by an ethnicity's representation in the government coalition, or by an ethnicity having lost or gained power. Second, the effect is not due to confounding geographic or climatic differences between an ethnic group and the ethnicities forming the government in a specific year. The estimates remain large and significant when controlling for ethnic differentials in geographic location, elevation, ruggedness, caloric suitability, mean temperature, and average yearly precipitation. Third, the results are not systematically driven by an ethnic group's conflict behavior prior to a government change. Taken together, these findings show that the estimates reflect the effect of cultural distance, and not of other confounding factors.

Next, I validate my findings using a tightly-controlled empirical exercise on a restricted sample of ethnic groups. My baseline specification does not control for unobserved time-varying, ethnic-specific shocks. If these occur simultaneously to a change in government and are correlated with an ethnic group's decision to rebel, they could confound my results. To address this issue, I restrict the analysis to the subset of ethnicities that were split by country borders during the Scramble for Africa and are therefore simultaneously exposed to different governments in different countries. While this subsample is arguably peculiar due to the documented long-lasting effects of ethnic partitioning on conflict (Michalopoulos and Papaioannou [2016]), focusing on this offers some advantage for the purpose of identification.

I use a triple difference-in-differences (DID) design that exploits within-ethnicity variation in linguistic distance to the government resulting from (i) a group's concurrent exposure to multiple governments in multiple countries and (ii) changes in the ethnic identity of the government in some countries but not others. Combined, these features allow me to enrich the baseline specification with the inclusion of ethnicity-year fixed effects.[6] Reassuringly, the results from the triple DID analysis are very similar to the ones from my baseline specification. A one standard deviation increase in cultural distance to the government increases the prevalence of conflict over government power by 0.46 standard deviations. A one standard deviation increase in cultural distance, conversely, reduces conflict over territory by 0.22 standard deviations, although this effect is only marginally significant. In line with the original hypothesis, this suggests that conflict over territory tends to be more likely between culturally close groups that share similar preferences.

[6]This approach is analogous to the one employed by Dickens [2018]. Michalopoulos and Papaioannou [2014] were the first exploiting the partition of African ethnicities in multiple countries as a source of within-ethnicity variation.

One remaining potential concern with the analysis is the presence of omitted factors affecting both cultural distance and conflict. For instance, historical conflict between two ethnicities might have a direct impact on the probability of current conflict between them, and at the same time, might have affected their linguistic distance. I address these endogeneity concerns using a novel instrumental variables (IV) approach. To instrument for cultural distance, I exploit the Bantu expansion, a massive pre-historical migration that changed language and culture in some parts of sub-Saharan Africa but not others.[7] To my knowledge, this is the first paper exploiting this prehistoric cultural shock as a source of exogenous variation in cultural distance between today's African ethnicities.

I construct and assign to each ethnicity a *Bantu index*, which captures the extent to which ethnic homelands were exposed to the route of Bantu migration.[8] I then instrument for cultural distance between two ethnic groups using the absolute difference in their Bantu index. The idea behind this instrument is the following. Groups whose homelands were highly exposed to the Bantu migration route inherited Bantu culture, and should be culturally distant to those whose culture remained unaffected. Conversely, two groups with a similarly high or similarly low Bantu exposure should be culturally close to each other, because they either both inherited Bantu culture, or they both kept their pre-existing one.[9] Consistent with this idea, the first stage documents a strong positive association between the absolute distance in the Bantu index and cultural distance. In the second stage, the effect of cultural distance on conflict over power becomes larger in magnitude when compared to the OLS estimates. A one standard deviation increase in cultural distance increases the prevalence of conflict by 0.86 standard deviations.[10] In the case of conflict over territory, the negative coefficient reaches statistical significance in some specifications. A one standard deviation increase in cultural distance decreases the prevalence of conflict over territory by 0.11 standard deviations.

In a last step, I explore factors that may explain why cultural differences trigger conflict over the control of the central government. I show that cultural differences are associated with diverging preferences over public policies. In particular, using individual-level data from seven rounds of the Afrobarometer survey for more than 80,000 individuals in 27 African countries, I find that respondents are more likely to oppose a wide range of current government policies if

[7]Bantu farmers imported their culture capital into some of the territories originally inhabited by Pygmy and Khoisan hunter-gatherers. Archaeological evidence, despite being inconclusive, suggests that Pygmy people inhabited Central Africa, and Khoisan people inhabited sub-equatorial Africa (Diamond [1997]). Both were hunter-gatherer societies, unlike the Bantu people, who were farmers. In gradually expanding from their homeland in Cameroon, Bantu people followed emerging savannah corridors, which appeared in the rainforest due an exogenous climatic shock and then disappeared once climate re-adjusted (Grollemund *et al.* [2015]).

[8]To generate the Bantu index, I merge contemporary ethnicities to their ancestral homelands in the Murdock map, the most comprehensive ethnolinguistic map of Africa prior to European contact, and overlay ethnic groups' homelands on the historical migration route of the Bantu reconstructed by Grollemund *et al.* [2015].

[9]The assumption here is that, within contemporary countries, pre-Bantu groups were on average culturally similar to each other, and culturally dissimilar to the Bantu. This is partly supported by archaeological evidence showing that pre-Bantu societies shared the same mode of subsistence, i.e., hunting and gathering, while Bantu people were predominantly farmers (Diamond [1997]).

[10]This suggests that the original coefficients were biased downwards. This would be consistent, for instance, with the following: past conflict between two ethnicities is positively associated with contemporary conflict, and, at the same time, reduced two groups' linguistic distance through genocides eliminating linguistically distant subgroups.

they belong to ethnicities that are culturally distant to the ethnic groups in power at the time of the survey. A one standard deviation increase in cultural distance increases discontent with government performance by 0.44 standard deviations.

This paper contributes to several strands of literature. First, it adds to the vast literature on civil war (see Blattman and Miguel [2010] for a review). Among studies focusing on ethnic conflict, some have highlighted factors that make certain ethnicities more likely to experience conflict, like segmentary lineage organization (Moscona *et al.* [2020]), inter-personal diversity (Arbatlı *et al.* [2020]), and ethnic partitioning (Michalopoulos and Papaioannou [2016]). However, even after acknowledging these different propensities to experience conflict, what drives ethnic groups to fight against other specific ethnicities remains unclear. By focusing on cultural distance between ethnic groups and governments, I relate ethnicities to their potential opponents and show that the prevalence of conflict is a function of the characteristics of both combatants, and not only of the characteristics of one side. Two studies adopt a similar approach. Guarnieri and Tur-Prats [2020] focus on the intensive margin of violence, and show that cultural distance in gender norms between ethnic belligerents increase the use of conflict-related sexual violence. Guariso and Rogall [2017] show that income inequality between an ethnic group and the leading group in a country increases the likelihood of ethnic conflict between them. Whereas their focus is on the economic drivers of conflict, my paper explores the role of deep-rooted cultural determinants.

By uncovering that cultural distance increases conflict over government power but not conflict over territory, this paper highlights the importance of considering the issue over which combatants fight when studying the determinants of conflict. In investigating the economic roots of local violence across Africa, McGuirk and Burke [2020] also distinguish between two types of conflict, and find that this distinction matters. They show that, in food-producing areas, the same positive global food price shock reduces conflict over the control of territory (what they call "factor conflict"), but increases conflict over the appropriation of surplus ("output conflict"). Campante *et al.* [2019] show that, in relatively non-democratic countries, conflict over government power is more likely to occur close to the capital city and that this does not hold for conflict over territory.

This study is closely related to theoretical and empirical work on ethnic diversity and conflict, summarized in the next section. Empirically, this literature has mostly focused on the relationship between diversity indices and conflict at the country level (Esteban and Ray [2011] and Esteban *et al.* [2012]). This country-level approach, however, remains silent on the group-specific heterogeneity in the decision to fight. By moving the analysis to the ethnicity-level, I unpack the country-level associations and identify precisely which ethnicities engage in conflict, and why.

More broadly, this paper adds to the literature on the consequences of ethnic diversity. In addition to studies focusing on conflict, others have analyzed a large range of economic outcomes (see Alesina and La Ferrara [2005] for a review), adopting as unit of analysis either countries, cities, or grid cells (see, for instance, Montalvo and Reynal-Querol [2017]). My study is part of

a new line of work that keeps the aggregate level of diversity constant and focuses on distances between entities within aggregate units. While I study ethnicities within countries, Gomes [2020] examines individuals within geographical radii, and finds that children of mothers who are linguistically distant from their neighbors have worse health outcomes.

Finally, I contribute to the literature exploring the determinants of ethnolinguistic diversity. Michalopoulos [2012] finds that contemporary ethnic diversity is rooted into geographic variability. Work by Ashraf and Galor [2013] shows that genetic diversity, determined during the prehistoric migration of humans out of Africa, is a fundamental determinant of ethnic heterogeneity within countries today. Galor *et al.* [2018] explore the precise geographic roots of specific cultural and linguistic traits like the structure of the future tense, sex-based grammatical gender, and politeness distinctions. A recent contribution by Dickens [2020] studies linguistic distance between neighboring groups. This study finds that ethnicities separated across geographic regions with high variation in land productivity are more similar than ethnicities separated across more homogeneous regions, due to higher levels of historical trade. I speak to this literature by linking the prehistoric Bantu migration to cultural distance between ethnic groups today.

The remainder of the paper is structured as follows. Section 2 describes the conceptual framework linking cultural distance to inter-group conflict. Section 3 outlines the data sources and the procedure I adopted to construct the dataset. Section 4 describes the empirical strategy, and Section 5 shows the main results together with a battery of robustness tests. Sections 6 and 7 present the difference-in-differences and the instrumental variables estimation strategies, respectively, and the corresponding results. Section 8 discusses potential channels, and Section 9 concludes.

3.2 Conceptual Framework

The link between diversity and conflict has been debated by a large strand of interdisciplinary literature. This debate originates from the so-called primordialist view of conflict, according to which dissimilarities between groups spur conflict.[11] An anthropological formulation of this view posits that a society formed by culturally divergent groups can only be sustained through a political order in which one of the cultural groups politically dominates the others (Smith [1965]). In turn, this political structure inevitably generates dissensus, given that all groups within a jurisdiction are bound to the public policies provided by a culturally distant dominant group. According to Smith's [1965] theory of cultural pluralism, it is this *cultural dissensus* that generates ethnic conflict in pluralist societies. Yet, cultural differences might also impede conflict, by "focusing the ambitions of various groups on alternative sources of gratification, thereby preventing them from impinging on each other" (Horowitz [2000], p. 138). Taken together, these opposing views already hint at the potentially ambiguous relationship between cultural differences and conflict.

[11] This view is often juxtaposed to the so-called instrumentalist view, which sees ethnic conflict as a result of grievances and inequality between groups, rather than the results of different ethnic identities.

Literature in economics has recently shed further theoretical and empirical light on these issues. In their study of inter-state conflict, Spolaore and Wacziarg [2016b] provide empirical evidence in support of Horowitz's [2000] conjecture. They find that inter-state conflict is more likely between similar populations, because they tend to share preferences over the same type of resources or territories—the typical bone of contention in international conflict.

In analyzing the relationship between alternative diversity measures at the country level and the equilibrium level of conflict, work by Esteban and Ray [2011] and Esteban *et al.* [2012] highlights that distinguishing between conflict over private goods and conflict over public goods is essential for understanding the interplay between diversity and conflict. Theoretically and empirically, they show that a country's severity of conflict can be approximated by a weighted average of three country-level measures of diversity: a Gini coefficient, an index of ethnic fractionalization, and a measure of ethnic polarization. The weights of each of these measures depend on a country-level measure of "publicness of the prize".

To elucidate this concept, Esteban and Ray [2011] discuss the following examples. In a situation in which groups fight for the control of an excludable private good, "primordial" inter-group distance—best captured by a measure of polarization—does not constitute a relevant explanation for this conflict. The only explanatory factor will be a country's distribution of group sizes, best captured by an index of fractionalization. At the other extreme, when a prize is fully public, differences in group preferences over public goods will emerge. This type of conflict will be better explained by indexes of polarization and Gini coefficients, since they capture the degree of inter-group distances within a country.

Along similar lines, Spolaore and Wacziarg's [2017] theoretical framework highlights that the impact of cultural distance on conflict depends on whether groups are fighting over the control of public goods or rival goods. Conflict over the control of public goods—e.g., public policies that all must share within a jurisdiction—is more likely between culturally distant groups that hold different preferences. This explanation is entirely consistent with Esteban and Ray's [2011] theory. However, while in Esteban and Ray [2011] distance does not matter for conflict over excludable private goods, Spolaore and Wacziarg's [2017] framework claims that conflict over rival goods—e.g., territories and resources—will be more likely among culturally close groups that share similar preferences. I bring this hypothesis to the data and, for the first time, simultaneously test these two predictions at the ethnic group level in the context of ethnic civil conflict.

3.3 Data and Descriptive Statistics

3.3.1 Sources and Dataset Construction

To examine the relationship between cultural distance and conflict, I assemble a unique dataset measuring how culturally distant an ethnic group is from the government of a country in a specific year, which groups engage in violent conflict at a certain point in time, and whether

they are fighting for government power or for exclusive control of a territory. I furthermore add information on climatic and geographical features, as well as on ancestral ethnic settlements of contemporary ethnic groups and their exposure to the route of the Bantu expansion. In this section, I describe the procedure I adopted to construct the dataset.

Ethnic groups. The first step consists of retrieving information on ethnic groups forming a country's population. To this end, I exploit the Ethnic Power Relations (EPR) Dataset Family (Wucherpfennig *et al.* [2012]). For each country of the world, EPR lists all politically relevant ethnic groups between 1946-2017 and their access to government power. A group is defined as *politically relevant* if one political actor claims to represent the interests of a group in the national political arena, or if group members are systematically and intentionally discriminated against. EPR codes the degree to which each ethnic group's representatives hold executive-level state power in each year, which ranges from total control of the government to political discrimination.[12] Each group falls into one of the following categories, in decreasing order of power: monopolist, dominant, senior partner, junior partner, self-exclusion, powerless, discriminated.[13] Based on this data, I construct a panel that follows each ethnic group over time.

Ethnic civil conflicts. To construct the dependent variable, I use the UCDP/PRIO Armed Conflict Dataset (Harbom *et al.* [2008]; Allansson *et al.* [2017]), which includes information on civil conflicts and the actors involved in them between 1946 and 2019. The UCDP/PRIO dataset defines civil conflicts as "armed conflicts between the government of a state and one or more internal opposition groups that cause at least 25 battle-related deaths within a year". I focus on *ethnic* civil conflicts in Africa fought between 1961 and 2017,[14] those where armed groups "explicitly pursue ethno-nationalistic aims, motivations, and interests; and recruit fighters and forge alliances on the basis of ethnic affiliation" (Wimmer *et al.* [2009]).

UCDP/PRIO provides information on the issue over which rebel groups fight. First, a conflict can be about *government*, i.e., about the type of political system, the replacement of the central government, or the change of its composition. Second, a conflict can be about *territory*. While groups fighting for government intend to impose their authority over others, groups fighting for territory do not want to exercise power over other groups, but strive for exclusive control of a specific territory for own settlement, local resource use, or, in extreme cases, secession. In the first case, the nature of the bone of contention is public, while in the latter case, the good combatants are fighting over is rival and excludable.

[12] Whenever government changes happen in the same year as a conflict, EPR reports the power relations that were in place before the conflict outbreak.

[13] Additional categories include state collapse and irrelevant. I exclude the (rare) instances of state collapse from the analysis as, in those years, there is no government; I exclude irrelevant groups for the analysis as, for those, there is no information on their settlements.

[14] I restrict the analysis to this time period as most African countries gained independence in the 60s, and climatic data is available only starting from 1961. I exclude the most recent year because the EPR data is available only until 2017.

Examples of struggles over state power include the case of the Liberians United for Reconciliation and Democracy (LURD) (2000-2003) against the government led by Charles Taylor, or the Tigray People's Liberation Front (TPLF) in Ethiopia (1976-1986) fighting against the hegemony of the Amharan government. Conflicts fought over territory include the one between the Movement of Democratic Forces of Casamance (MFDC) and the government of Senegal for the control of the Casamance region (1990-2011). Another example is the rebellion of the Liberation of the Enclave of Cabinda-Renewal (FLEC-R) and the Armed Forces of Cabinda (FLEC-FAC) against the government of Angola over the control of the Cabinda region (1991-2017). This category also includes cases where rebel groups seek control over specific resources linked to a certain territory, such as the Niger Delta People's Volunteer Force (NDPVF) fighting for controlling the petroleum resources of the delta region in Nigeria (2004).

Rebels' ethnic identity. To assign rebel groups an ethnic identity, I exploit the ACD2EPR dataset (Vogt *et al.* [2015]). This dataset assigns to each rebel group one or more ethnic groups in the EPR dataset. To continue some of the examples above, members of the Mandingo and Krahn ethnic groups formed the LURD rebel group in Liberia; the Tigry ethnic group formed the TPLF rebel group in Ethiopia; and the Ijaw group formed the NDPVF in Nigeria. Next, I merge this information to the panel of ethnicities, where each ethnic group constitutes a *potential rebel*, and construct a binary variable that is equal to 1 if a potential rebel is involved in a conflict against the central government in a certain year, and zero otherwise. All groups within a country can be potential rebels, apart from *monopolists* or *dominant* ones. By definition, these groups cannot rebel against themselves, because they have the exclusive control of the government.[15] Whenever an ethnicity becomes dominant or monopolist, it drops out of the panel.[16] In a robustness test, I re-run the analysis excluding these ethnicities, and show that the results remain unchanged.

Governments' ethnic identity. To assign an ethnic identity to the government in a specific year, I exploit information provided by the EPR dataset, which indicates which ethnic groups hold executive government power in a country at a certain point in time. EPR takes into consideration where executive power is exercised when coding ethnicities' access to it. This includes the presidency, the cabinet, and senior posts in the administration, including the army, or the ruling party leadership in one-party states. In some cases, government power is held exclusively by one ethnic group, that EPR classifies as either *monopolist* or *dominant*. In other cases, the government results from a coalition of ethnic groups, which can take either the role of *junior* or *senior partner*. Whether a group is classified as senior ot junior partner depends on the group's absolute influence in the executive, measured by the number and importance of the positions controlled by group members.

[15]Indeed, there are no instances in which a rebel group involved in a conflict is associated with a dominant or monopolist ethnicity.

[16]Note that these cases are quite rare. Only 14 ethnic groups temporarily exit the sample of potential rebels because they become monopolists or dominant.

Cultural distance. To measure cultural distance between each potential rebel and the government, I use linguistic distance, as is standard in the literature. The source of information on languages is the EPR Ethnic Dimensions (EPR-ED) dataset. EPR-ED assigns up to three languages to each EPR ethnic group, indicating the three largest language segments spoken by group members in descending order and their relative size.[17] The advantage of this dataset is that it provides a nuanced description of the linguistic (and therefore, cultural) identity of an ethnicity also when the latter is not homogeneous. Therefore, my measure of cultural distance incorporates this intra-group heterogeneity.

I merge languages to linguistic trees in the Ethnologue database (Lewis *et al.* [2020]). For each language, the Ethnologue provides a classification starting with the language family (e.g. Afro Asiatic, Nilo-Saharan, Creole), followed by "nodes", i.e., the branching points of the linguistic tree, and ending with the language itself. To compute linguistic distance between each potential rebel and the government, I employ a three-steps procedure.

First, I calculate the distance between each pair of languages (x and y) as follows:

$$d_{xy} = 1 - \left(\frac{\#\text{ of common nodes between } x \text{ and } y}{\frac{1}{2}(\#\text{ of nodes of language } x + \#\text{ of nodes of language } y)} \right)^{\lambda}$$

This measure, called cladistic distance, is the most frequently used in the literature (Fearon and Laitin [2003]). Languages originating from different families have no nodes in common, and their distance will be equal to 1. The parameter λ ranges between 0 and 1, and is used to attribute higher weight to earlier common nodes, as early separations in the language tree are likely to signify larger cultural divergence on average than later separations. I follow Fearon and Laitin [2003] and assign to λ a value of 0.5.

Second, I calculate linguistic distance between potential rebels (r) and each ethnic group at the government (g_i):

$$LD_{rg_i} = \sum_{x=1}^{3} \sum_{y=1}^{3} (w_{rx} \times w_{g_i y} \times d_{xy}) \tag{3.1}$$

where w_{rx} and $w_{g_i y}$ are the fraction of population speaking language x in group r and language y in group g_i, respectively.

Third, I compute linguistic distance between each potential rebel and the government:

$$LD^W = \sum_{i=1}^{N} p_{g_i} \times LD_{rg_i} \tag{3.2}$$

where N is the total number of ethnicities forming the government, and p_{g_i} is a weight reflecting the position of power of group g_i in the government. When the government is composed only by one dominant or monopolist ethnic group, LD^W equals LD_{rg_i}. In the case of a government

[17] These refer to indigenous African languages, not those imported by colonizers.

coalition, I assign a higher weight to linguistic distance between each potential rebel and the senior partner.[18] Senior partners will receive double the weight of each junior partner.[19]

Alternatively, I use an unweighted version of the just described linguistic distance measure:

$$LD^{UW} = \sum_{i=1}^{N} \frac{LD_{rg_i}}{N} \tag{3.3}$$

Ancestral settlements. I merge EPR ethnicities to ethnic groups in the Murdock Ethnographic Atlas (EA). This allows me to assign ethnic groups to the location of their ancestral homelands prior to European contact, which was recorded by Murdock. I merge contemporary ethnic groups to pre-colonial ethnic groups and respective settlements through the Linking Ethnic Data for Africa (LEDA) R-package (Müller-Crepon *et al.* [Forthcoming]), which systematically links ethnic categories across datasets exploiting the Ethnologue linguistic tree.[20] The definition of ethnic group in the EA is more fine-grained than the one in the EPR dataset. Therefore, one ethnic group in EPR is generally matched to multiple groups in the Murdock map.

Bantu expansion route. To estimate the exposure of each ethnic group to the Bantu expansion, I digitize the Bantu expansion route reconstructed by Grollemund *et al.* [2015] (see Figure 3.4, left). I then overlay the path onto the Murdock map. I leverage ethnic groups' ancestral homelands to calculate the exposure to the Bantu expansion for the following reasons. First, EPR provides ethnic settlements for contemporary ethnic groups, which are likely the result of recent phenomena endogenous to conflict, as well as of modern-day country borders imposed artificially by colonial powers. Second, the geographic locations of ethnic groups in EPR capture groups' regional presence and contemporary settlements, rather than homelands as in the Ethnographic Atlas.

Geographic and climatic controls. The GeoEPR dataset provides polygons describing each ethnic group's geographic location, allowing overlaps between different groups' settlements.[21] To add information on geographic and climatic characteristics to each settlement, I combine GeoEPR with data on elevation, temperature, and precipitation at the grid level provided by Worldclim (Fick and Hijmans [2017]). Furthermore, to add information on the potential agricultural output, I use the Caloric Suitability Index (CSI) (pre-1500 average) from Galor and

[18] Rebels may be intentionally fighting against one specific group in the coalition, or against the government as a whole. Since I do not have systematic information on this issue, I consider all ethnic groups forming the government.

[19] In the example illustrated in Table 3.1, in the distance between Langi/Bakongo and the government, p_{g_i}=0.5 when g_i=Mbochi (the senior partner), and p_{g_i}=0.25 when g_i=Bateke or Kouyou (the junior partners).

[20] For additional details on the methodology, see Müller-Crepon *et al.* [Forthcoming]. For details on how I use the LEDA R-package, see Appendix B.

[21] GeoEPR does not provide the location of groups classified as living exclusively in urban areas. To these, I assign the coordinates of the country's capital city. These constitute 2% of ethnicities in my sample.

Table 3.1: *Data Extract and Illustration of the Identifying Variation*

country	year	potential rebel	Government: ethnic group 1 [*status*]	ethnic group 2 [*status*]	ethnic group 3 [*status*]	ethnic group 4 [*status*]	linguistic distance	conflict over power	conflict over territory
Congo	1995	Lari/Bakongo	Bembe [*Senior partner*]	Bateke [*Junior partner*]	Kouyou [*Junior partner*]	Vili [*Junior partner*]	0.098	0	0
Congo	1996	Lari/Bakongo	Bembe [*Senior partner*]	Bateke [*Junior partner*]	Kouyou [*Junior partner*]	Vili [*Junior partner*]	0.098	0	0
Congo	1997	Lari/Bakongo	Bembe [*Senior partner*]	Bateke [*Junior partner*]	Kouyou [*Junior partner*]	Vili [*Junior partner*]	0.098	0	0
Congo	1998	Lari/Bakongo	Mbochi (proper) [*Senior partner*]	Bateke [*Junior partner*]	Kouyou [*Junior partner*]		0.163	1	0
Congo	1999	Lari/Bakongo	Mbochi (proper) [*Senior partner*]	Bateke [*Junior partner*]	Kouyou [*Junior partner*]		0.163	1	0
Congo	2000	Lari/Bakongo	Mbochi (proper) [*Senior partner*]	Bateke [*Junior partner*]	Kouyou [*Junior partner*]		0.163	0	0

Notes: The table illustrates a data extract for the potential rebel Lari/Bakongo in Congo between 1995 and 2000. The identifying variation is within-ethnicity changes in cultural distance resulting from government changes. In this case, the change in government occurred in 1998, Lari/Bakongo experienced an increase in cultural distance, and started rebelling against the government.

Özak [2016]. This measure reflects the potential caloric yield of a territory using a combination of geographical and climatic data unaffected by human activity.

Opinions on government performance. I merge ethnicity-level data to individual-level information on opinions about government performance collected in 7 waves of the Afrobarometer Survey (Afrobarometer Data [1999-2017]). I link EPR groups to the self-reported linguistic affiliation of individuals in the Afrobarometer through the LEDA R-package (Müller-Crepon *et al.* [Forthcoming]).

Figure 3.1 and Table 3.1 illustrate the structure of my data and the identifying variation with an example from Congo. First, as shown in Figure 3.1, there are six politically active ethnic groups: Lari/Bakongo, Kouyou, Mbochi (proper), Bateke, Bemba, Vili. The map illustrates where each ethnic group is located within the country. I follow each of these ethnicities over time, and track their involvement in conflict as well as their cultural distance to the government. Table 3.1 reports a data extract for the ethnic group Lari/Bakongo between 1995 and 2000. The within-ethnicity variation in cultural distance I exploit is driven by government changes: in this case, Lari/Bakongo experienced an increase in linguistic distance in 1998, following the Mbochi group replacing the Bembe as a senior partner in the coalition, and the Vili group dropping out of power. After this change, Lari/Bakongo started engaging in conflict against the government.

3.3.2 Summary Statistics

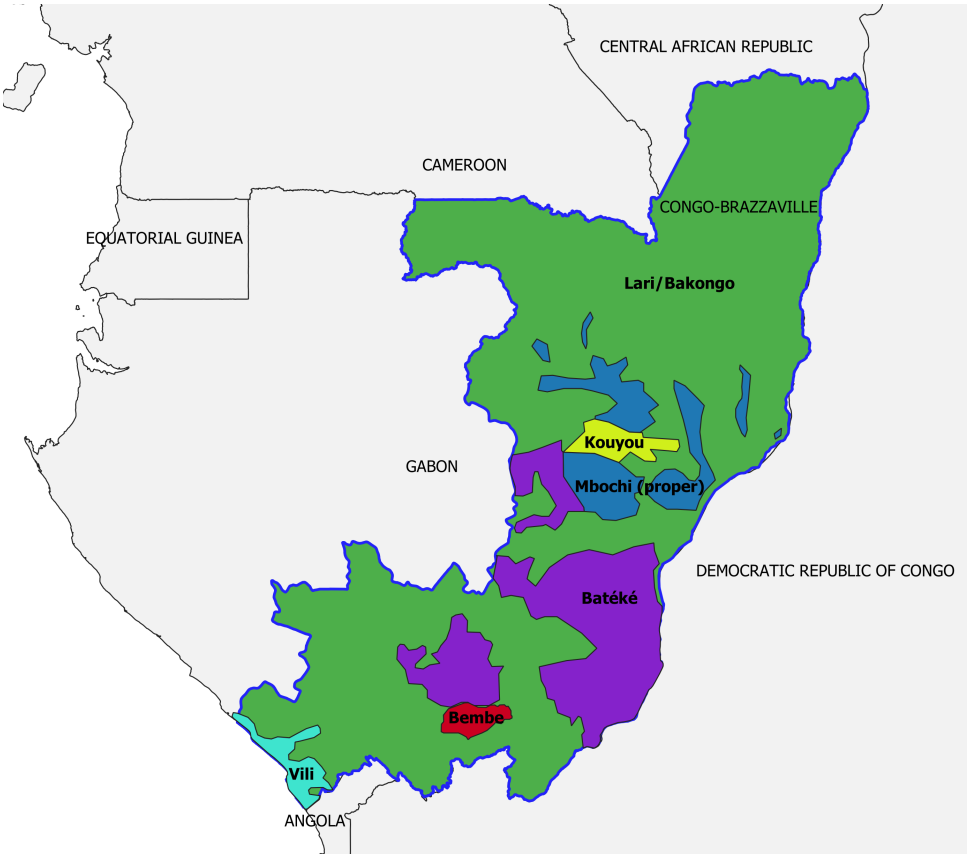
The resulting dataset includes 236 ethnic groups in 43 African countries over a period of 57 years (1961-2017) (N=9,827).[22] The source of within-ethnicity variation in cultural distance stems from 138 changes in the ethnic composition of the government. Table C.1 reports the number of changes experienced by each country in the sample. Ten countries out of 43 did not experience any change in the ethnic composition of the government in the period considered. The largest part of the variation comes from West Africa, with countries such as Niger, Nigeria, and Congo experiencing the highest number of government transitions.

Table C.2 reports descriptive statistics of the main variables used in the analysis. In Africa as a whole, 29 countries experienced at least one conflict over the period considered, with an overall prevalence of ethnic conflict of 0.073. Of the 236 potential rebels in the sample, 73 became rebels at least once, and 60 distinct governments were involved in an ethnic civil conflict at least once. 34 percent of the overall conflict prevalence is explained by conflicts fought for territory.

Within countries in Africa, linguistic distance between potential rebels and the government is on average 0.40-0.45, depending on the distance type considered. Potential rebels and the government are culturally close in Malawi, with an average linguistic distance of 0.06, while potential rebels are culturally very distant from the government in Liberia (0.90 on average).

[22] I exclude countries where there is only one ethnic group (Burkina Faso, Cabo Verde, Lesotho, Somalia, and Tunisia), and countries where there is only one potential rebel (Rwanda, Madagascar, Algeria). I do not include Sao Tome and Principe and Seychelles since they are not part of the EPR dataset family.

Figure 3.1: *Ethnic Groups and Ethnic Settlements in Congo*



Notes: The figure illustrates the six politically relevant ethnic groups (Lari/Bakongo, Kouyou, Mbochi (proper). Bateke, Bemba, Vili) in Congo-Brazzaville and their settlements. Source: Ethnic Power Relations (EPR) Core (Wucherpfennig *et al.* [2012]) and GEO-EPR (Wucherpfennig *et al.* [2011]).

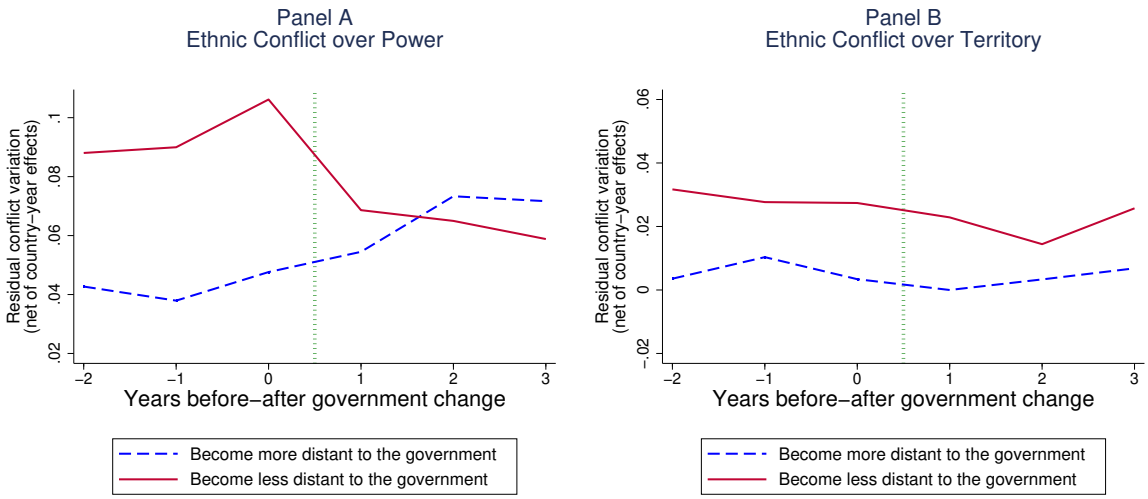
The Liberian example speaks to the accuracy of linguistic distance in capturing cultural differences. The Americo-Liberians were the group in power over almost the whole period considered. This group originates from free-born and formerly enslaved African Americans who emigrated in the 19th century and became the founders of Liberia. They imported African-American and Caribbean culture, and are therefore culturally different from the other Liberian groups (Gio, Mano, Mandingo, Kran (Guere), and Indigenous peoples).

3.4 Empirical Strategy

The objective of the empirical analysis is to investigate whether changes in cultural distance to the government have an effect on ethnic groups' involvement in ethnic civil conflict. To illustrate the idea behind the empirical strategy and get a first sense of how the prevalence of conflict reacts to government changes, in Figure 3.2 I plot the average residual variation in conflict over power and conflict over territory—net of country-year fixed effects—before and after a government change. I do so separately for two subsamples of ethnicities: groups that become

more culturally distant to the government; and groups that become less culturally distant to the government following a change in leadership. Panel A shows that a group's involvement in conflict over power reacts to cultural distance: groups are more likely to rebel once they become more culturally distant to the government, and less likely once they become culturally closer. Conversely, Panel B illustrates that groups' involvement in conflict over territory does not respond to cultural distance. Importantly for identification, both panels indicate that the two subgroups of ethnicities do not display different trends in conflict involvement prior to a government change.

Figure 3.2: *Ethnic Conflict Before and After a Government Change*



Notes: The figure plots the average residual variation in conflict over power (left) and conflict over territory (right) net of country-year fixed effects for two subsamples: ethnic groups that become more culturally distant to the government following a government change (blue dashed), and ethnic groups that become less culturally distant to the government following a government change (red solid).

To measure the impact of changes in cultural distance on conflict in a regression framework, I estimate the following linear probability model:

$$\text{Conflict}_{rct} = \lambda_{c,t} + \zeta_r + \theta_r t + \beta \text{LD}_{rct} + \Gamma \text{G}_{rct} + \Omega \text{C}_{rct} + \epsilon_{rct} \tag{3.4}$$

where the dependent variable is an indicator that takes value 1 if there is a rebel group of ethnicity r fighting the government of country c in year t. LD_{rct} indicates linguistic distance between ethnic group r and the government of country c in year t.

$\lambda_{c,t}$ denotes a full set of country-year fixed effects. These capture time-invariant characteristics at the country-level—e.g., colonial history or geography—that might make some countries overall more or less prone to conflict than others. Furthermore, they account for time-specific shocks common to all ethnic groups in a country and for a specific government's characteristics. These include whether the government represents a large versus a small share of the population; whether it is a coalition of groups or formed by a single dominant one; whether it is a bellicose or peaceful government; the quality and type of its public policies; the strength of the army; or

whether a certain government is established as a result of a conflict and thus is more vulnerable to retaliation from opposing groups. Finally, country-year fixed effects allow to isolate the effect of cultural distance on conflict from the effect of a country's aggregate level of fractionalization and polarization.

Crucially, my dataset also allows me to include a full set of ethnicity fixed effects (ζ_r), as well as ethnicity-specific year trends ($\theta_r t$). By adding ethnicity fixed effects, I control for any time-invariant characteristics specific to an ethnic group that might affect its propensity to experience conflict. These include all ethnic traits that have been associated with conflict by the literature like social structure (Moscona *et al.* [2020]), within-group inter-personal diversity (Arbatlı *et al.* [2020]), or ethnic partitioning (Michalopoulos and Papaioannou [2016]). Moreover, these fixed effects ensure that the variation I exploit is within-ethnicity variation in cultural distance resulting from government changes (see example in Table 3.1). By controlling for an ethnicity-specific year trend, I can furthermore account for dimensions at the ethnicity level that change linearly over time, and allow each ethnic group to follow different trends.[23]

Finally, I control for a set of geographic and climatic controls. G_{rct} is a vector of differences between potential rebels and the government: geodesic distance, distance in elevation, ruggedness, and caloric suitability index. Dickens [2020] shows that geographical differences are an important determinant of linguistic distance, as neighboring groups that share homogeneous environments are more likely to be linguistically closer to each other. These findings speak to the importance of this vector of controls when seeking to isolate the effect of cultural distance from the effect of geography. I also include a set of differences in precipitation and mean temperature (C_{rct}) to account for the effect of rainfall inequality on ethnic conflict uncovered by Guariso and Rogall [2017]. I also control for time-variant climatic variables (temperature and precipitation) specific to each potential rebel to account for the potential confounding effect of climatic shocks, which are important determinants of conflict in the African continent according to Harari and La Ferrara [2018].[24]

3.5 Main Results

Table 3.2 reports estimates from the linear probability model described in equation 3.4, using as outcome variable a binary measure of conflict, either fought over government power or territory. Column 1 shows that within a country and a year, the larger the linguistic distance between an ethnic group and the government, the more likely it is that the group engages in conflict against

[23]Note that this specification is equivalent to a difference-in-differences strategy where treatment is the change (positive or negative) in cultural distance stemming from government changes. I compare groups that experience a positive change in cultural distance to those that experience a negative change. The inclusion of unit-specific time trends allows me to relax the common trends assumption typical of DID designs. See a similar application in Pfeifer and Reutter [2020]. Another assumption for a causal interpretation of my results is that treatment effects have to be homogeneous across ethnic groups within a country, as well as over time (Goodman-Bacon [2018]).

[24]Given the dynamic nature of the GeoEPR dataset, ethnic settlements may vary over time, making geographic characteristics also time-variant. However, since ethnicity fixed effects absorb the variation almost entirely, I do not control for potential rebels' elevation, ruggedness, and CSI to avoid collinearity issues.

the ethnicities in power. In this specification, without the inclusion of ethnicity fixed effects and ethnicity time trends, a one standard deviation increase in linguistic distance increases conflict prevalence by 1.3 percentage points (0.05 standard deviations). This specification, which exploits within-country cross-ethnicity variation in cultural distance, does not account for plausibly relevant ethnic-specific confounding factors.

Table 3.2: *Cultural Distance and Ethnic Civil Conflict*

	Dependent variable: ethnic conflict					
	(1)	(2)	(3)	(4)	(5)	(6)
Linguistic distance^{W}	0.013***	0.060***	0.063***	0.072***		
	(0.004)	(0.014)	(0.017)	(0.019)		
Linguistic distance^{UW}					0.051***	0.061***
					(0.015)	(0.022)
Mean of dep. var.	0.073	0.073	0.073	0.073	0.073	0.073
Country-year fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects		yes	yes	yes	yes	yes
Ethnicity year trend		yes	yes	yes	yes	yes
Geographic controls			yes	yes		yes
Climatic controls				yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.239	0.524	0.525	0.522	0.521	0.522

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government and 0 otherwise. *Linguistic distance*^{W} captures linguistic distance between each potential rebel and the ethnic groups at the government, weighted by the position of power of each ethnic group in case the government is a coalition. *Linguistic distance*^{UW} is an unweighted average of linguistic distance between each potential rebel and each ethnic group at the government. The linguistic distance measures are standardized. Geographic controls include the logged geodesic distance between each ethnic group and the government, absolute distance in elevation, absolute distance in ruggedness, and absolute distance in the caloric suitability index (CSI). Climatic controls include yearly average precipitation, yearly average temperature, absolute distance in mean precipitation, and absolute distance in mean temperature to the ethnic groups in power. The sample is an unbalanced panel on 236 distinct ethnic groups in 43 African countries over 57 years (1961-2017). Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

To account for these, I add ethnicity fixed effects in column 2, and now exploit within-ethnicity variation in cultural distance originating from changes in the ethnic composition of the government. The coefficient remains positive and significant, and increases in magnitude. A one standard deviation increase in cultural distance increases conflict prevalence by 6 percentage points (0.23 standard deviations). This coefficient remains significant and similar in magnitude when adding geographic and climatic controls (columns 3 and 4) and when using an unweighted measure of linguistic distance to the government (column 5).

In Table 3.3 I test my main hypothesis, i.e., whether these associations mask heterogeneous effects depending on what combatants are fighting over. I find that the just described effects

are entirely driven by conflicts fought over government power. In my preferred specification in column 2 of Table 3.3, a one standard deviation increase in cultural distance increases conflict over power by 7.5 percentage points (0.36 standard deviations). When comparing results for the weighted measure of cultural distance in columns 1 and 2 to those for the unweighted measure in columns 3 and 4, one sees that the effect of cultural distance is slightly larger when accounting for the power structure in the government coalition. A senior partner will have larger decision-making power over the provision of public policies compared to a junior partner. Therefore, the larger effects resulting from a weighted measure of cultural distance are in line with my hypothesis: an ethnicity rebels to gain power when disagreeing with the mix of public policies provided by a culturally distant government. In Section 3.8 I will provide direct evidence on this specific mechanism.

Conversely, I do not find that increases in cultural distance to the government trigger conflict over territory (columns 5 to 8). The negative sign of the insignificant coefficients suggests that, if anything, the association reverses: a potential rebel is less likely to fight for territory when experiencing an increase in cultural distance to the government.

3.5.1 Leads and Lags and Anticipatory Effects

In this section, I explore the persistence of the effect of cultural distance on conflict, and test for anticipatory effects. Table 3.4 presents the results of these exercises. Column 1 reproduces the baseline estimation for linguistic distance in t for reference. To explore whether the effect persists, column 2 reports the estimate for a lagged measure of linguistic distance. The positive and significant coefficient suggests that the effect of a government change on conflict persists, but tends to fade away: the magnitude of the coefficient in column 2 almost halves, when compared to the one in column 1.[25] When including both contemporaneous and lagged linguistic distance simultaneously in column 3, only the contemporaneous measure remains positive and significant, suggesting that ethnic groups' reaction to a culturally distant (or close) government is immediate.

One concern could be that ethnic groups becoming more distant to the government were already following different trends in conflict involvement prior to a leadership change, when compared to ethnic groups becoming more culturally close. If this was the case, my estimates could be capturing pre-existing differences in trends between these two groups of ethnicities, and would not be the result of a change in cultural distance. While the visual inspection of the temporal variation in conflict in Figure 3.2 already suggests that this is not the case, I nevertheless formally test the common trends assumption in a regression framework.

In column 4 of Table 3.4, I add a lead measure of linguistic distance. Should there be any pre-trends in the ethnic group(s) that become more distant to the government compared to those that become less distant after a leadership change, then this lead measure of linguistic distance should be estimated significantly different from zero. The coefficient is insignificant, pointing

[25] See Table C.16 in the Appendix for the full set of estimates using the lagged measure of cultural distance as main explanatory variable.

Table 3.3: *Cultural Distance and Ethnic Civil Conflict over Government Power and Territory*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.065***	0.075***			-0.004	-0.004		
	(0.015)	(0.017)			(0.008)	(0.011)		
Linguistic distance^{UW}			0.061***	0.072***			-0.009	-0.011
			(0.014)	(0.016)			(0.011)	(0.014)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.491	0.491	0.490	0.492	0.587	0.587	0.587	0.587

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns (1 -4)) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns (5 -8)). For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

to absence of anticipatory effects and suggesting that the common trends assumption, crucial for identification, is satisfied. In columns 6-10, I report the same sequence of estimates with conflict over territory as dependent variable and, again, I do not find evidence of anticipatory effects (column 10). In the following subsection, I present a set of robustness tests.

3.5.2 Robustness Tests

Since the variation driving my results stems from government changes, a concern with my estimation strategy could be that ethnic groups enter or exit power in t as a result of their involvement in conflict (or lack thereof) in $t-1$. Given that ethnic conflict tends to be persistent over time, my results could be driven by a group's involvement in conflict prior to government transitions, and thus have nothing to do with changes in cultural distance. To assess this, I re-run the main specification controlling for a binary variable that is equal to 1 if a potential rebel was involved in a conflict over the previous year. Results in columns 1 to 4 in Table 3.5 show that lagged conflict is a powerful predictor of current conflict, but that the coefficients on linguistic distance—albeit smaller in magnitude—remain positive and significant in the case of conflict over government, and negative and insignificant in the case of conflict over territory.[26] This suggests that my estimates are not an outcome of potential rebels' conflict involvement prior to a government change.

A potential explanation for my results could also be related to an ethnicity's representation in the government coalition. When becoming part of a coalition, an ethnicity's cultural distance to the government mechanically decreases. At the same time, this could prevent an ethnic group to rebel for reasons not necessarily related to cultural distance (e.g., the desire not to hurt members of the same group). Conversely, an increase in cultural distance might coincide with an ethnicity losing power, and this could trigger conflict aimed at regaining a position in the government. While I discuss these as potential mechanisms for my results in Section 3.8, I here perform three initial tests to verify whether this constitutes a threat to the interpretation of my results.

First, in Table 3.6, I test whether the effect of cultural distance on conflict occurs over and above the effect of being represented in the government. To assess this, I run a horse-race between cultural distance and a binary variable that equals 1 if a potential rebel is represented in the government coalition and 0 otherwise. Second, in Tables C.7 and C.8, I control for whether an ethnicity has lost or gained power in the previous year, respectively. As expected, these variables are associated with conflict over government power: an ethnicity is less likely to

[26] One caveat with this specification is that the inclusion of a lagged dependent variable in a fixed effects model might generate Nickell bias (Nickell [1981]). A solution to this problem could be to instrument the first lag of the dependent variable with the second lag. I report results of this exercise in Table C.6 in the Appendix. However, note that this solution requires arguably strong assumptions (i.e., that the second lag be uncorrelated with the first-differenced residuals). In line with Guryan's [2001] arguments, I expect that the estimates in Table 3.3 constitute an upper bound of the true effects, while those in Table 3.5 a lower bound. This would be consistent with treatment (i.e., an increase in cultural distance) being *negatively* selected on lagged conflict (see Appendix 1 in Guryan [2001]), meaning that a *decrease* in cultural distance might be the result of successful past conflict over government power.

Table 3.4: *Cultural Distance and Ethnic Civil Conflict: Leads and Lags, Anticipatory Effects*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance$_{t-1}$		0.039**	-0.037	-0.041		-0.002	0.005	0.010
		(0.018)	(0.025)	(0.026)		(0.012)	(0.009)	(0.007)
Linguistic distance$_{t}$	0.075***		0.109***	0.079**	-0.004		-0.010	-0.004
	(0.017)		(0.028)	(0.036)	(0.011)		(0.010)	(0.005)
Linguistic distance$_{t+1}$				0.032				-0.005
				(0.029)				(0.008)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,565	9,565	9,305	9,827	9,565	9,565	9,305
Adjusted R-squared	0.491	0.500	0.503	0.510	0.587	0.595	0.595	0.626

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns (1 -4)) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns (5 -8)). *Linguistic distance* captures weighted linguistic distance between each potential rebel and the ethnic groups at the government, either in period t-1, t, or t+1. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table 3.5: *Cultural Distance and Ethnic Civil Conflict: Controlling for Lagged Conflict*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.034***	0.036**			-0.004	-0.004		
	(0.010)	(0.014)			(0.007)	(0.010)		
Linguistic distance^{UW}			0.030***	0.033**			-0.009	-0.010
			(0.011)	(0.015)			(0.010)	(0.012)
Conflict$_{t-1}$	0.572***	0.571***	0.572***	0.571***	0.405***	0.405***	0.405***	0.404***
	(0.081)	(0.082)	(0.081)	(0.082)	(0.115)	(0.115)	(0.115)	(0.114)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,565	9,565	9,565	9,565	9,565	9,565	9,565	9,565
Adjusted R-squared	0.672	0.672	0.672	0.672	0.662	0.662	0.662	0.662

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns (1 -4)) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns (5 -8)). Conflict$_{t-1}$ is a binary variable that is equal to 1 if an ethnic group was involved in a conflict over power (columns (1-4)) or over territory (columns (5 -8)) in the previous year. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table 3.6: *Cultural Distance and Ethnic Civil Conflict: Controlling for Representation in Government Coalition*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.040**	0.055***			-0.008	-0.008		
	(0.015)	(0.020)			(0.008)	(0.011)		
Linguistic distance^{UW}			0.037**	0.051**			-0.015	-0.015
			(0.017)	(0.020)			(0.012)	(0.014)
Ethnic group is part of coalition	-0.075**	-0.081**	-0.076**	-0.082**	-0.013	-0.013	-0.016	-0.017
	(0.033)	(0.038)	(0.034)	(0.039)	(0.012)	(0.015)	(0.012)	(0.015)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.501	0.501	0.500	0.501	0.593	0.593	0.593	0.593

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns (1 -4)) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns (5 -8)). *Ethnic group is part of coalition* is equal to 1 if an ethnicity is represented in the government coalition in a certain year and 0 otherwise. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

rebel when holding a position in the government coalition, and more likely to rebel after having lost power. However, while the magnitude of the cultural distance estimates slightly decreases, the significance of the main results is unaffected by the inclusion of these controls. Third, I construct an alternative measure of cultural distance, which excludes the potential rebel from the distance computation if the latter is part of the government coalition. Results are reported in Table C.9. Again, all previous conclusions remain valid.

Next, I check the robustness of my results to alternative specifications, estimation samples and clustering of standard errors. First, I show that results are similar when excluding ethnic-specific time trends (Table C.12). In Figure C.1, I check for outliers by re-running my estimations dropping one country at a time. I also restrict the analysis to the 29 countries that experienced at least one conflict between 1961 and 2017 (Table C.10) and to the sample of 33 countries that experienced at least one government change (Table C.11). In Tables C.13 and C.15 I re-run the analysis with an alternative level of clustering of standard errors, by country, country-year or by ethnicity, respectively. My results are robust to all these tests.

Finally, I run the main specification on a balanced panel of 207 ethnicities (out of the original 236) that I observe uninterruptedly over the sample period. The original dataset is an unbalanced panel due to some missing country-years (e.g., Sierra Leone during the state collapse between 1993-2001), or due to ethnicities exiting the sample while holding monopoly or dominant power in the government.[27] Estimates on this subsample of ethnicities, reported in Table C.17, are similar to those from the full sample.

3.6 Triple Difference-in-Differences Approach

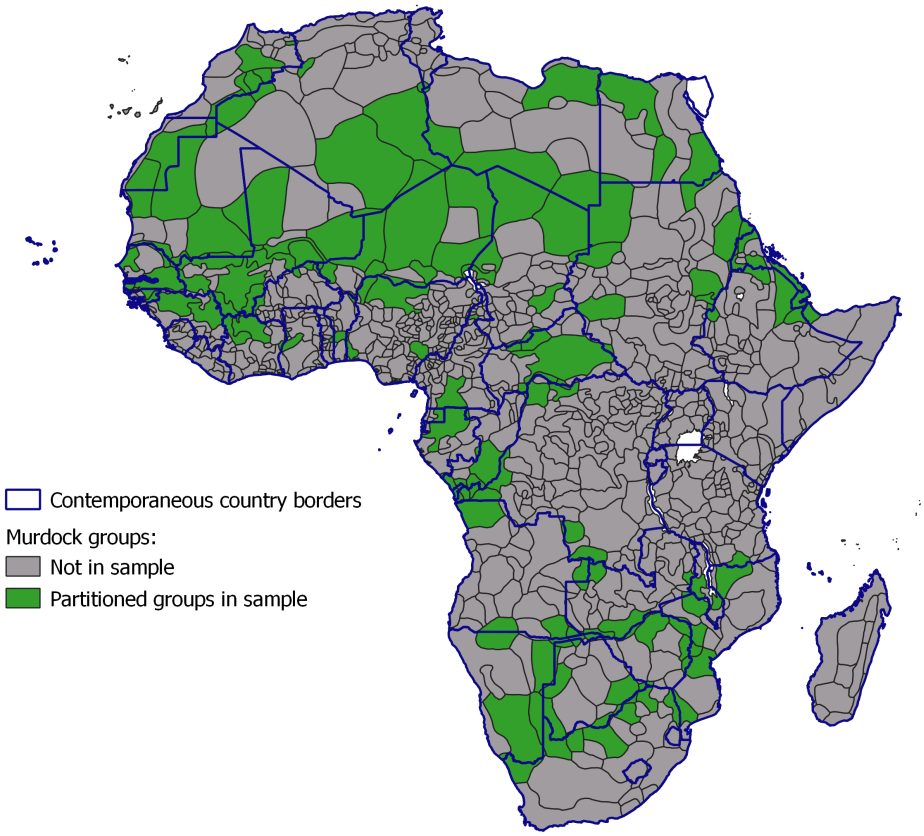
So far, my estimation strategy did not allow to account for unobserved time-varying ethnic-specific shocks. If these occurred simultaneously to a change in government and were correlated to an ethnic group's decision to rebel, they could confound my results. To address this, I re-run the analysis on a subset of ethnic groups that were split by country borders during the Scramble for Africa. This strategy, similar to the one employed by Michalopoulos and Papaioannou [2014] and analogous to the one adopted by Dickens [2018], exploits the fact that the same ethnicity is simultaneously exposed to different governments in different countries. As the ethnic identity of governments changes in some countries but not others, the quasi-random allocation of borders provides an exogenous source of within-group variation in cultural distance to the government.[28] As cultural distance of partitioned groups varies over time and between countries, this variation allows me to estimate a triple difference-in-differences specification and thus control for a full set of ethnicity-year fixed effects.[29]

[27] As mentioned in Section 3.3, these groups cannot be potential rebels as, by definition, they cannot be linked to rebel groups fighting the government.

[28] For additional evidence and details on the randomness in the allocation of colonial boundaries, see the discussion in Michalopoulos and Papaioannou [2016] and in Dickens [2018].

[29] The categories underlying the triple DID are: culturally distant/culturally close; treated (e.g., experiencing a government change)/non-treated; before/after a government change.

Figure 3.3: *Murdock Ethnic Groups Split by Country Borders*



Notes: The figure displays the ethnic groups in the Murdock Ethnographic Atlas that were split by country borders and that I include in the difference-in-differences analysis. I do not include groups that I cannot successfully merge to at least two groups in the EPR dataset belonging to different countries.

While this estimation strategy constitutes a more tightly-controlled empirical exercise, it has the disadvantage of restricting the sample to a peculiar set of ethnicities, i.e., those that were split into multiple countries. As Michalopoulos and Papaioannou [2016] show, partitioned groups are considerably more likely to experience political violence, ethnic wars, and government-led discrimination when compared to non partitioned ethnicities. For this reason, one should interpret the results presented in this section as a robustness exercise in support of the baseline findings.

I merge contemporary ethnicities to the Murdock Atlas, which contains information on ethnic settlements prior to European contact and therefore provides the most reliable source to identify groups that were partitioned during the Scramble for Africa.[30] The map in Figure 3.3 displays the partitioned ethnicities that I successfully merged with the EPR ethnic groups. This restricted sample includes 96 distinct Murdock ethnic groups partitioned across 32 African

[30]This is also the data source employed by Michalopoulos and Papaioannou [2016], who were the first exploiting partitioned groups for identification.

countries.[31] Table C.3 reports summary statistics. Compared to the full African sample, conflict prevalence is slightly higher, which is consistent with Michalopoulos and Papaioannou's [2016] findings showing that partitioned groups are more likely to experience conflict compared to non-partitioned ones. On average, linguistic distance to the government is also higher in this restricted sample.

I estimate the following empirical model:

$$\text{Conflict}_{rct} = \eta_{c,r} + \lambda_{c,t} + \zeta_{r,t} + \beta \text{LD}_{rct} + \Gamma \text{G}_{rct} + \Omega \text{C}_{rct} + \epsilon_{rct} \tag{3.5}$$

The dependent variable is a measure of conflict for ethnic group r in country c and year t. The main independent variable, LD_{rct} is a measure of linguistic distance to the government in period t.

As in my previous specifications, I add a full set of country-year fixed effects ($\lambda_{c,t}$), as well as a set of ethnicity-country fixed effects ($\eta_{c,r}$). Moreover, since the same potential rebel is present in multiple countries, this specification controls for a full set of ethnicity-year fixed effects ($\zeta_{r,t}$), and thus accounts for any time-variant ethnic-specific shocks. G_{rct} and C_{rct} indicate the same set of geographic and climatic controls outlined in section 3.4

3.6.1 Results

Table 3.7 reports the results. Columns 1 to 4 show estimates for conflict over government power, using both the weighted and the unweighted measure of cultural distance. In column 2, a one standard deviation increase in cultural distance to the government leads to a 10.2 percentage points (or 0.46 standard deviations) increase in disputes over power. The effect is large when compared to the mean prevalence of 0.051 in the sample considered, and remains quantitatively similar to the effect uncovered using the whole African sample.

When turning to conflict over territory in columns 5 to 8, the association reverses. Albeit significant only at the ten percent level, the estimate in column 6 indicates that a one standard deviation increase in cultural distance to the government decreases conflict over territory by 3.6 percentage points (or 0.22 standard deviations), a sizable effect that compares to a 0.026 conflict prevalence in the sample.

As in the estimating equation in the whole African sample, identification rests on the assumption that governments are not systematically elected because of the conflict behavior of ethnic groups prior to an election. In Table C.19, I find that the estimates remain similar, albeit less precisely estimated, when controlling for a lagged dependent variable. In Table C.20, I show that my main coefficients are robust to the inclusion of a binary variable indicating whether an ethnic groups is part of the government coalition or not as a control. Finally, in Table C.18, I report estimates for leads and lags of cultural distance and, similar to the results for the whole African sample, I do not find evidence of anticipatory effects.

[31] Note that my merging procedure is successful for 41 percent of the ethnicities that Michalopoulos and Papaioannou [2016] classify as partitioned.

Table 3.7: *Difference in Differences: Exploiting Ethnic Groups Split by Country Borders*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.075*	0.102**			-0.025*	-0.036*		
	(0.038)	(0.039)			(0.014)	(0.019)		
Linguistic distance^{UW}			0.084**	0.120***			-0.027*	-0.035*
			(0.041)	(0.038)			(0.014)	(0.018)
Mean of dep. var.	0.051	0.051	0.051	0.051	0.026	0.026	0.026	0.026
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	7,874	7,874	7,874	7,874	7,874	7,874	7,874	7,874
Adjusted R-squared	0.858	0.860	0.858	0.862	0.794	0.799	0.794	0.799

Notes: the unit of observation is a Murdock ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns (1 -4)) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns (5 -8)). For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 32 African countries, 57 years (1961-2017), and 96 distinct Murdock ethnic groups. Following Dickens [2018], standard errors are clustered at the ethnic group-country level in parenthesis (corresponding to 201 clusters). Inference is similar when clustering two-way by country and year, or by ethnicity and year. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

3.7 Instrumental Variables Approach

A remaining concern with my strategy is that the measure of cultural distance might be endogenous to conflict. If there are omitted factors that affect linguistic distance between groups and contemporary conflict, then the estimates presented in the previous section are biased. An example is historical conflict. Old antagonisms between groups might have affected linguistic distance and, at the same time, can have a direct effect on conflict today.[32]

This is particularly a concern in my setting, where not all members of an ethnic group speak the same language. Recall that the linguistic distance measure is a weighted average of all pairwise linguistic distances, where weights reflect the percentage of group members speaking each language. Therefore, any source of endogeneity could operate through two channels: there could be omitted factors affecting both conflict and (i) how the linguistic tree of two languages evolved, or (ii) the relative importance of a language within an ethnicity's population.

The direction of the potential bias is ex-ante unclear. Suppose past conflict between two ethnicities is positively associated with current conflict. A history of violence might also have reduced two ethnicities' linguistic distance over time, e.g., due to genocides eliminating linguistically distant subgroups. If this was the case, then the main coefficients on conflict over government power would be biased downwards. On the other hand, if past conflict widened linguistic distance by reinforcing linguistic enclaves or ethnic boundaries, then my coefficients would be biased upwards.

3.7.1 The Instrument

To instrument for cultural distance between groups, I exploit the Bantu expansion, a natural experiment unique to African history. 5,000 years ago, a climatic shock generated a temporary loss of rainforest in Central Africa. Through increased seasonality of the monsoon, a lowering of the sea surface temperature in the Guinean Gulf, and less rainfall, the shock favored the emergence of savanna corridors (Bostoen *et al.* [2015]). After the climatic crisis ended, the new savanna environments disappeared, supplanted by rainforest.

Archaeological, anthropological, and linguistic evidence suggests that the temporary opening up of parts of the—beforehand impenetrable—rainforest facilitated the expansion of the Bantu, a tribe residing in Cameroon, throughout sub-Saharan Africa (see Bostoen *et al.* [2013] and Bostoen *et al.* [2015] for a summary of the interdisciplinary literature). The exact reason for why Bantu people started migrating is largely debated. One hypothesis is that the climate-driven opening of the forest gave hunters access to "naive" animal populations that were previously trapped in the forest and then became suddenly available (Bostoen *et al.* [2013]). However, there is by now consensus among scholars on how the *path* of the Bantu migration was exogenously shaped by climatic events, a crucial feature for my instrumental variables analysis (Grollemund *et al.* [2015], Bostoen *et al.* [2015]).

[32] For evidence on how past conflict is correlated with current conflict in Africa, see Besley and Reynal-Querol [2014].

How is this massive prehistoric migration related to contemporary cultural distances between ethnic groups? The Bantu expansion has been defined as the most important linguistic, cultural and demographic process in Late Holocene Africa, and one of the most influential cultural events of its kind (Robbeets and Savelyev [2017]). Bantu farmers, in gradually expanding from their homeland in Northwestern region of Cameroon, spread their new language and culture throughout sub-Saharan Africa, assimilating or displacing earlier inhabitants of the regions they crossed, i.e., Pygmy and Khoisan hunter-gatherers (Diamond [1997]). Today, one out of three Africans is fluent in at least one of the approximately 500 existing languages belonging to the Bantu family.

Crucially, Bantu people did not move and settle everywhere in Sub-Saharan Africa, and this is exactly the idea behind my instrument. The Bantu expansion acted as a prehistoric "cultural shock"—in turn exogenously triggered by a climatic crisis—which affected the cultural capital of some areas but not others. Within today's countries, some territories—which then became ethnic homelands—happened to be crossed by the route of the Bantu expansion, while others did not. This is demonstrated, for instance, by the existence of pre-Bantu hunter-gatherers populations such as the Pygmies in Central Africa, the Hadza in Tanzania, and Khoisan people in southern Africa. The route of the Bantu expansion did not cross their ethnic homelands, and for this reason, these groups remain today culturally and linguistically distinct.

I exploit this unique event in African prehistory to construct a novel instrument for cultural distance between ethnic groups. Groups whose homelands were highly exposed to the Bantu migration route have inherited Bantu culture, and therefore should be culturally distant to those that remained unaffected. Instead, groups with a similarly high or low exposure to the Bantu expansion should be culturally close to each other, because they either both inherited Bantu culture, or kept their pre-existing one.[33] Based on this idea, I use the absolute difference in two groups' exposure to the Bantu expansion as an instrument for cultural distance.

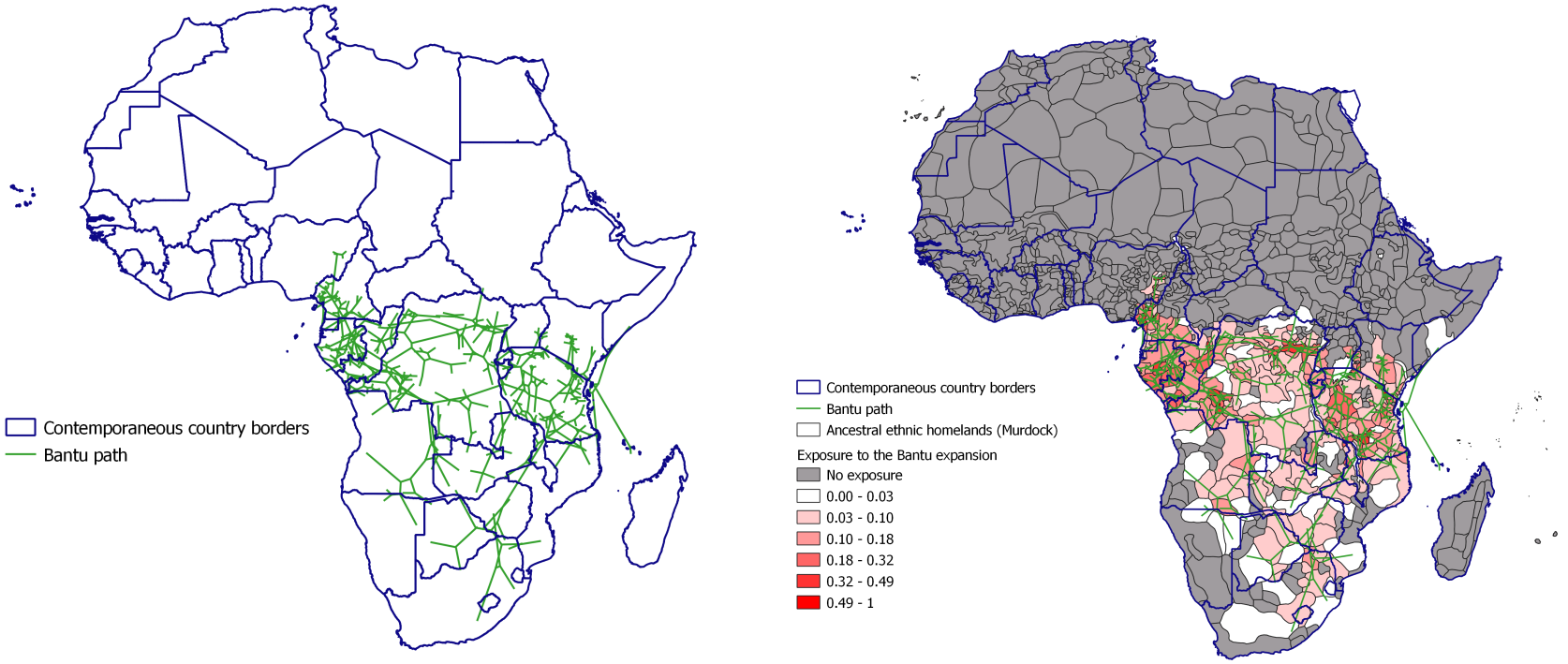
To obtain the instrument, I start by constructing a measure of exposure to the Bantu expansion for each ethnic group's homeland. I merge contemporary ethnicities to their ancestral homelands in the Murdock map, the most ancient record of ethnic settlements in Africa. I then overlay ethnic groups' ancestral homelands on the historical migration route of the Bantu reconstructed by Grollemund *et al.* [2015].[34] The left panel of Figure 3.4 illustrates the route of the Bantu migration and contemporaneous country borders. For each ethnic group in the Murdock map, I construct a Bantu Index capturing the intensity of exposure to the Bantu migration route. I calculate the length of the path crossing each ethnic homeland, divide it by the size of the settlement, and then normalize this index to range between 0 and 1:

$$\text{Bantu Index}_{EA} = \frac{\text{Path Length}}{\text{Homeland Area}} \tag{3.6}$$

[33]The assumption here is that, within contemporary countries, pre-Bantu groups were on average culturally similar to each other, and culturally dissimilar to the Bantu. This is partly supported by archaeological evidence showing that pre-Bantu societies shared the same mode of subsistence, i.e., hunting and gathering, while Bantu people were predominantly farmers (Diamond [1997]).

[34]See Appendix B for details on how Grollemund *et al.* [2015] reconstructed the route.

Figure 3.4: *The Path of the Bantu Expansion and the Murdock Map*



Notes: The figure on the left shows the historical migration route of the Bantu tribe reconstructed by Grollemund et al. (2015) and contemporaneous country borders. See Appendix B for additional details on the route reconstruction. The figure on the right overlays historical migration route with the map of pre-colonial ethnic settlements constructed by Murdock. Different colors denote the intensity of exposure of each ethnic group to the Bantu expansion, computed as the the length of the path crossing an ethnic homeland divided by the size of the ethnic settlement, all normalized to range between 0 and 1.

Figure 3.4 (right) shows the distribution of the Bantu index across Africa. There is considerable within-country, cross-ethnicity variation in ethnic groups' exposure to the Bantu expansion. Since each EPR group is associated to multiple groups in the Ethnographic Atlas, I proceed as follows to obtain an index at the EPR-group level:

$$\text{Bantu Index}_{EPR} = \sum_{i=1}^{N} \alpha_i \text{Bantu Index}_{EA_i} \tag{3.7}$$

where α_i is a weight reflecting the accuracy of the match between each Murdock group and the EPR group.[35]

Next, I construct the instrument, i.e., the absolute distance in the Bantu Index between a potential rebel and the government. To this end, I follow the same procedure I adopted for the linguistic distance measure. First, I compute the Bantu distance between a potential rebel (r) and each ethnic group forming the government (g_i):

$$\text{BD}_{rg_i} = |\text{Bantu Index}_r - \text{Bantu Index}_{g_i}| \tag{3.8}$$

Second, I compute a weighted Bantu distance between each potential rebel and the government:

$$BD^W = \sum_{i=1}^{N} p_{g_i} \times BD_{rg_i} \tag{3.9}$$

where N denotes the total number of ethnicities forming the government and p_{g_i} is a weight reflecting the position of power of group g_i in the coalition. Alternatively, I construct the unweighted version:

$$BD^{UW} = \sum_{i=1}^{N} \frac{BD_{rg_i}}{N} \tag{3.10}$$

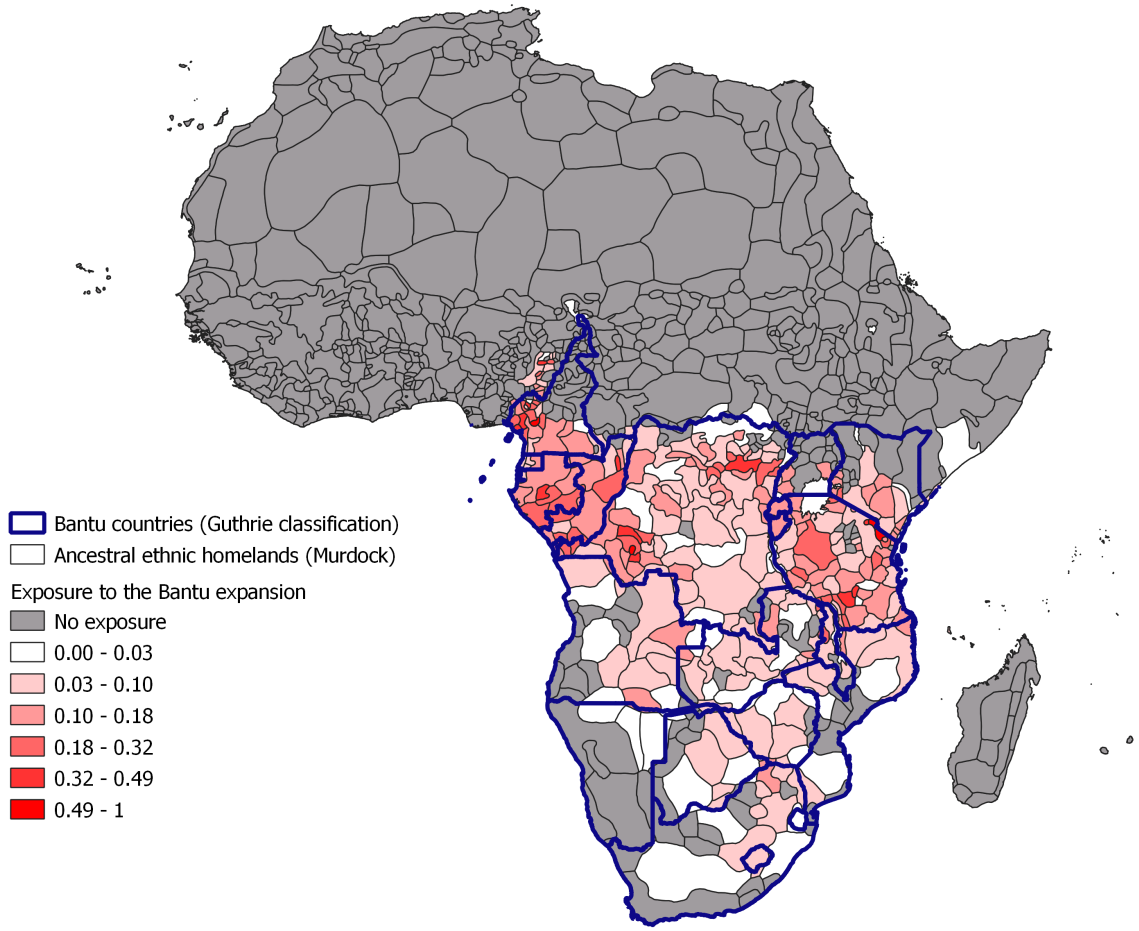
3.7.2 Empirical Strategy

I use a two-stage least-square (2SLS) procedure to estimate equation 3.4. In the first stage, I estimate the effect of the distance in the exposure to the Bantu expansion route on linguistic distance between ethnic groups today:

$$\text{LD}_{rct} = \lambda_{c,t} + \zeta_r + \theta_r t + \beta \text{BD}_{rct} + \Gamma \text{G}_{rct} + \Omega \text{C}_{rct} + u_{rct} \tag{3.11}$$

[35] The accuracy is given by the percentage of nodes in the EPR-group linguistic tree covered by the Murdock-group linguistic tree. $\sum_{i=1}^{N} \alpha_i = 1$. If the match is equally accurate between each Murdock group, then $\text{Bantu Index}_{EPR} = \sum_{i=1}^{N} \frac{\text{Bantu Index}_{EA_i}}{N}$. See Appendix C.2 for details on the LEDA package, the merging procedure, and the accuracy of the match.

Figure 3.5: *The Bantu Index and Countries in the Bantu Region*



Notes: The figure displays the index of exposure to the Bantu expansion based on the historical migration route reconstructed by Grollemund et al. (2015), as well as contemporaneous country borders of the 17 countries classified as belonging to the Bantu region according to the Guthrie classification. These include: Angola, Botswana, Burundi, Cameroon, Congo, Congo DRC, Equatorial Guinea, Gabon, Kenya, Malawi, Mozambique, Rwanda, South Africa, Tanzania, Uganda, Zambia, Zimbabwe.

I run the IV analysis for the subset of 17 sub-Saharan countries where the Bantu tribe migrated, and that are classified as Bantu countries according to the Guthrie classification of languages (see Figure 3.5).[36]

Table C.4 reports summary statistics for this subsample of 86 ethnic groups (N=3,628). The identifying variation stems from 41 government changes. In the Bantu region, the prevalence of conflict is slightly lower than in the continent as a whole (0.060) and only 13 percent of it is explained by conflicts fought for territory. Linguistic distance between groups is also, on average, smaller (0.257).

[36]The countries covered by Bantu languages according to the Guthrie classification are: Angola, Botswana, Burundi, Cameroon, Congo, Congo DRC, Equatorial Guinea, Gabon, Kenya, Malawi, Mozambique, Rwanda, South Africa, Tanzania, Uganda, Zambia, Zimbabwe.

3.7.3 Results

Table 3.8 reports results from the instrumental variables analysis, which confirm the linear probability model estimates.[37] In Panel A, columns 1 and 2 and columns 5 and 6 reports OLS estimates for conflict over government and conflict over territory, respectively, for the subsample of ethnicities in the Bantu region. Results are consistent with the estimates in the full African continent. The association between cultural distance and conflict over government is positive and significant, while the coefficients on conflict over territory are negative (and, only in one case, significant).

Panel C reports the first-stage results of equation 3.11 and shows how a larger difference in the exposure to the Bantu expansion is positively associated with cultural distance, measured through linguistic distance. The Kleibergen-Paap F-statistic indicates that the instrument is a powerful predictor of cultural distance. In Panel B (columns 3 and 4), I find that there is a positive reduced-form effect of Bantu distance on conflict over government power, which is reassuring for the validity of the instrument. A one standard deviation increase in Bantu distance increases conflict prevalence by 6.5-7 percentage points. Columns 3 and 4 in Panel A report the 2SLS result, which are consistent with the OLS estimates of columns 1 and 2. In all specifications, both for the weighted and unweighted measures of cultural distance, a one standard deviation increase in linguistic distance significantly increases the prevalence of conflict over power. The magnitude of the coefficients more than doubles when compared to the OLS specification. The original coefficients were biased downwards, which is consistent, for instance, with past conflict between two ethnicities being positively associated with contemporary conflict, and, at the same time, with past conflict reducing two groups' linguistic distance through genocides eliminating linguistically distant subgroups.

When turning to conflict over territory, columns 7 and 8 in Panel B document a negative reduced-form effect of Bantu distance on conflict over territory. In Panel A, columns 7 and 8 report the 2SLS results. There is a negative effect of cultural distance on conflict over territory, but only for the unweighted version of the linguistic distance measure not accounting for the relative importance of ethnic groups in the government (column 8). In this specification, a one standard deviation increase in cultural distance to the government decreases conflict by one percentage point, which is a large effect when considering the mean conflict prevalence in the sample (0.008).[38]

These results are robust to controlling for lagged conflict (Table C.23) and to controlling for whether the potential rebel is present in the government coalition or not (Table C.24).

[37] Table C.22 in the Appendix reports estimates for overall conflict.

[38] A recent contribution by Lee *et al.* [2020] proposes a new test to assess the validity of the second-stage t-ratio inference when the first-stage F statistic is smaller than 104.7, as in this case. Table C.21 in the Appendix reports second-stage results with standard errors based on Lee *et al.*'s [2020] *tF* procedure and shows that inference considerations remain unchanged.

Table 3.8: *IV: Bantu Exposure, Cultural Distance and Ethnic Civil Conflict*

	Ethnic conflict over government				Ethnic conflict over territory			
Panel A	OLS		2SLS		OLS		2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance	0.072***	0.070***	0.191***	0.139***	-0.000	-0.002**	-0.008	-0.010**
	(0.016)	(0.012)	(0.023)	(0.025)	(0.001)	(0.001)	(0.005)	(0.004)
Mean of dep. var	0.052	0.052	0.052	0.052	0.008	0.008	0.008	0.008
Panel B			Reduced Form				Reduced Form	
Bantu distance			0.070***	0.065***			-0.003*	-0.004**
			(0.013)	(0.013)			(0.002)	(0.002)
Mean of dep. var			0.052	0.052			0.008	0.008
Adjusted R-squared			0.537	0.534			0.330	0.330
Panel C			First Stage				First Stage	
			Linguistic distance				Linguistic distance	
Bantu distance			0.365***	0.468***			0.365***	0.468***
			(0.083)	(0.056)			(0.083)	(0.056)
Kleibergen-Paap F-statistic			14.62	53.64			14.62	53.64
Distance type	w	uw	w	uw	w	uw	w	uw
Country-year FE	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity FE	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity trends	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls	yes	yes	yes	yes	yes	yes	yes	yes
Climatic controls	yes	yes	yes	yes	yes	yes	yes	yes
Observations	3,628	3,628	3,628	3,628	3,628	3,628	3,628	3,628

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting for government power (columns (1-4)) or for territory (columns (5-8)) and 0 otherwise. *Linguistic distance* captures linguistic distance between each potential rebel and the ethnic groups at the government, either weighted or unweighted. *Bantu distance* denotes the absolute difference in the exposure to the Bantu expansion between potential rebels and the government. Linguistic distance and Bantu distance are standardized. For a description of geographic controls, refer to the notes in Table 3.2. The sample includes 17 African countries in the Bantu region, 57 years (1961-2017), and 86 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis, and are adjusted for the low number of clusters using the number of countries. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

3.7.4 Instrument Validity

The validity of the IV estimates rests on the assumption that the differential exposure to the Bantu expansion affects conflict today only through its impact on cultural distance, conditional on the controls included in the regression. The first concern is that geographical differences between ancestral homelands might have determined the path of the Bantu migration and, at the same time, might also affect the likelihood of conflict between groups today. First, Grollemund *et al.* [2015] emphasize how the route of the expansion was mostly determined by emerging savannah corridors, which appeared in the rainforest due to a climatic shock and then disappeared once the climatic crisis was over. Despite this suggestive evidence of exogeneity in the way the path evolved, my preferred specification addresses this concern by controlling for a set of geographic and climatic controls.

Another concern could be that the Bantu, in gradually expanding, intentionally and systematically avoided pre-existing tribes with certain characteristics that make them more (or less) prone to conflict today. Suppose the Bantu avoided particularly bellicose tribes, and favored territories inhabited by more peaceful populations. Then my instrument would be systematically picking up differences in bellicosity. If these differences have a direct effect on conflict today, the exclusion restriction is violated.

I examine whether this example is at play in my setting by delving deeper in the reduced-form analysis. If the Bantu systematically avoided certain territories along unobservable characteristics correlated with contemporary conflict, then only being more (or less) affected to the Bantu expansion should have a reduced-form effect on conflict, but not both. I test this by splitting the absolute distance in the Bantu index into two: being *more* exposed to the Bantu expansion than the ethnicities at the government, and being *less* exposed.

I run an alternative reduced-form estimation using these two measures, instead of the absolute difference as in my main specification. As shown in Table C.25, I do not find evidence of a different reduced-form effect of being more versus less exposed to the Bantu expansion. Both coefficients hold the same sign, and the F statistics indicates that these coefficients are not different from each other. This suggests that what matters for conflict between ethnic groups is their *absolute* distance in the Bantu index—as conjectured in section 3.7.1—and not the fact of being more (or less) affected by the Bantu migration compared to the other group.

3.8 Channels

Why are ethnic groups more likely to fight over power—but not over territory—when the government is culturally distant from them? In this section, I analyze potential mechanisms and explore pathways that have been advanced by existing theoretical work on relatedness and conflict.

My finding that cultural distance matters only for conflict fought over the control of the central government is consistent with Spolaore and Wacziarg's [2017] and Esteban and Ray's

[2011] theoretical frameworks. Deep-rooted cultural differences shape diverging preferences over public policies, which trigger conflict over government power, but not over territory. The notion that cultural distance between ethnic groups generates diverging preferences over public policies has been often assumed by the literature on diversity and conflict, but has not been directly tested empirically.

To provide direct evidence of this mechanism, I exploit individual-level data from the Afrobarometer survey. The sample comprises individuals belonging to 95 different ethnic groups in 27 African countries surveyed between 1999 and 2017. Respondents report their opinions regarding the performance of their current government in matters related to the following broad policy domains: the economy, education, health, infrastructure, minorities, national security, and other social issues. Using PCA, I take the first component and group opinions about government performance into a single index ranging between 0-1, with 1 denoting the highest degree of disagreement. Table C.26 reports the results of the PCA for each survey round. The first principal component explains 34-61% of the common variance of the domains. Each domain always loads positively on the first principal component, suggesting that individuals who dislike government performance tend to do so along the whole range of listed public policies.

To establish whether cultural distance to the government comes with higher disagreement over public policies, I run the following specification:

$$\text{Opinion government}_{irct} = \lambda_{c,t} + \zeta_r + \theta_r t + \beta \text{LD}_{rct} + \Gamma \text{G}_{rct} + \Omega \text{C}_{rct} + \Phi \text{I}_i + \epsilon_{irct} \quad (3.12)$$

where $\text{Opinion government}_{irct}$ is a continuous variable that ranges between 0 and 1 and captures the extent to which individual i belonging to ethnic group r thinks that the government of country c in year t is performing fairly badly or very badly. LD_{rct} indicates linguistic distance between individual i's ethnic group r and government g in country c and year t. In addition to the set of fixed effects and ethnicity-level controls included in the baseline specification (see equation 3.4), I also add a set of individual-level controls I_i (a dummy for female, age, age squared, and a dummy for urban residents) and a survey round fixed effect.[39]

Results in Table 3.9 show that individuals are more likely to disagree with a wide range of public policies implemented by the government when they are more linguistically distant to the ethnic groups in power. One standard deviation increase in an individual's linguistic distance to the government increases the degree of opposition to government policies by 13-20 percentage points, which is a large effect, considering an average degree of disagreement of 56 percent in the sample (a 23-36 percent increase). Equivalent results arise when constructing the dependent variable taking the mean disagreement across all policy domains.

Another potential explanation for my results is related to Caselli and Coleman's [2013] theory of ethnic conflict. In their framework, language is an ethnic marker that allows groups in power to exclude non-members from enjoying a country's resources. For example, linguistic distance

[39] Note that each Afrobarometer round is conducted during different years in different countries.

facilitates marking non-group members and discriminating against them in the moment they apply for goods distributed through the central government, such as scarce places in higher education or jobs in the public sector. Ethnic distance should come with more conflict, because it is easier for a dominant group to mark and discriminate against ethnically distant groups.

According to this interpretation, it could be that discrimination—facilitated by linguistic distance—is what triggers conflict over power, and not a cultural mechanism. To shed light on this alternative explanation, I explore whether linguistic distance triggers conflict over power through discrimination. To this end, I run the following:

$$\text{Conflict}_{rct} = \lambda_{c,t} + \zeta_r + \theta_r t + \beta \text{LD}_{rct} + \kappa \text{S}_{rt} + \rho \text{S}_{rt} \times \text{LD}_{rct} + \Gamma \text{G}_{rct} + \Omega \text{C}_{rct} + \epsilon_{rct} \quad (3.13)$$

where S_{rt} indicates one of the three following binary variables: (i) equal to 1 if a potential rebel r is in power in year t and 0 otherwise; (ii) equal to 1 if a potential rebel r is powerless in year t and 0 otherwise; (iii) equal to 1 if a potential rebel r is discriminated in year t and 0 otherwise. κ captures the direct effect of a potential rebel's status on its conflict behavior at time t, and ρ—the coefficient on the interaction term—measures to what extent the effect of linguistic distance on conflict over government power interacts with an ethnic group's status (in power, powerless or discriminated) in the power-relations structure at time t.

Table 3.10 reports estimates of equation 3.13. Columns 1 to 3 display coefficients when S_{rt} is a binary variable that equals 1 if a potential rebel is in power (i.e., part of the government coalition), and zero otherwise. As already shown in the robustness tests in section 3.5.2, and as expected, being in power has a direct negative effect on conflict. However, the effect of linguistic distance on conflict does not seem to occur through an ethnic group's position of power in the government coalition: the coefficient on the interaction term ρ is small and insignificant, and the coefficient β remains positive and significant. Similarly, being powerless does not have a direct effect on conflict (see columns 4 to 6), nor does the interaction term with linguistic distance.

Finally, in columns 7 to 9 I directly test whether discrimination is what drives the effect of linguistic distance on conflict, consistently with the explanation outlined above. As expected, being discriminated has a strong positive direct effect on the likelihood that a potential rebel is fighting the government over power. However, the coefficient on linguistic distance remains large and significant, and the interaction term is insignificant. While I acknowledge the limitation of this empirical exercise due to the endogeneity of the discrimination measure, I take these results as additional suggestive evidence in favor of the cultural mechanism.

3.9 Conclusion

Conflict—and in particular, civil conflict fought along ethnic lines—is more prevalent in ethnolinguistically diverse societies. Yet, despite facing the same aggregate level of ethnolinguistic diversity, some ethnic groups rebel and others do not. This paper sheds light on why this is the case by studying the relationship between an ethnicity's cultural distance to the central

Table 3.9: *Cultural Distance and Opinions on Government Performance*

	Dependent variable: Thinks the government is performing badly					
	First principal component			Average number of items		
	(1)	(2)	(3)	(4)	(5)	(6)
Linguistic distance	0.207***	0.207***	0.134**	0.178***	0.178***	0.126**
	(0.050)	(0.050)	(0.054)	(0.050)	(0.050)	(0.053)
Mean of dep. var.	0.559	0.559	0.559	0.555	0.555	0.555
Country-year fixed effects	yes	yes	yes	yes	yes	yes
Survey round fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity year trends	yes	yes	yes	yes	yes	yes
Individual controls		yes	yes		yes	yes
Geographic controls			yes			yes
Climatic controls			yes			yes
Observations	83,482	83,482	83,482	90,481	90,481	90,481
Adjusted R-squared	0.179	0.179	0.180	0.186	0.187	0.187

Notes: the unit of observation is an individual belonging to one of 95 distinct ethnic groups in 27 African countries interviewed between 1999 and 2017 in 6 rounds of the Afrobarometer survey. The dependent variable is a binary variable capturing the extent to which the person thinks the government is performing very badly or fairly badly in handling the following matters: managing the economy, creating jobs, keeping prices stable, narrowing income gap, reducing crime, improving basic health services, addressing educational needs, improving water and sanitation services, ensuring enough to eat, fighting corruption, reducing conflict, combating malaria, combating HIV, maintaining roads and bridges, providing reliable electric supply, managing rivers, promoting equal rights/opportunities for women, addressing needs of youth, protecting rights and promoting opportunities for disabled. Columns 1-3 adopt as dependent variable the first principal component in a principal component analysis grouping all items separately for each survey round. The first principal component loads positively on all items. Columns 4-6 adopt as dependent variable the mean of all items. *Linguistic distance* captures the weighted linguistic distance between each potential rebel and the ethnic groups at the government. *Linguistic distance* is standardized. For a description of geographic controls, refer to the notes in Table 3.2. Individual controls include age, age squared, a dummy for female and a dummy for residence in a rural area. Standard errors clustered at the ethnicity-year level are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table 3.10: *The Role of Power Relations*

	Dependent variable: Ethnic conflict over government								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Linguistic distance	0.038**	0.052**	0.057**	0.074***	0.076***	0.084***	0.053***	0.063***	0.069***
	(0.017)	(0.020)	(0.023)	(0.020)	(0.020)	(0.020)	(0.016)	(0.017)	(0.018)
In power	-0.038**	-0.041**	-0.040**						
	(0.017)	(0.019)	(0.019)						
Powerless				-0.017	-0.018	-0.019			
				(0.012)	(0.012)	(0.012)			
Discriminated							0.080**	0.082**	0.082***
							(0.030)	(0.030)	(0.030)
Linguistic distance	-0.002	0.002	0.002	-0.007	-0.006	-0.006	-0.010	-0.011	-0.011
× Status	(0.014)	(0.016)	(0.016)	(0.007)	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)
Mean of dep. var.	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trends	yes	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes	yes		yes	yes		yes	yes
Climatic controls			yes			yes			yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.494	0.494	0.494	0.499	0.499	0.499	0.512	0.512	0.512

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government and 0 otherwise. *Linguistic distance* captures linguistic distance between each potential rebel and the ethnic groups at the government. *Status* is a binary variable which takes value 1 if an ethnic group is *in power* (columns 1-3), *powerless* (columns 4-6), or discriminated by the government (columns 7-9), and 0 otherwise. For a description of geographic controls, refer to the notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

government and its propensity to rebel against it. Consistently across estimation strategies, I find that a group's involvement in conflict increases with its cultural distance to the central government. However, this is true only for a specific type of conflict, i.e., one that is fought over the control of the central government. Ethnic rebellions over territory do not respond to changes in a group's distance to the government. If anything, larger distance comes with less conflict over territory and resources.

My paper relies on two key innovations. First, it draws a distinction between conflicts fought over government power and conflicts fought over territory. Second, it moves the analysis from the country level to the ethnicity level. Previous empirical analyses on diversity at the country level did not consider the ethnic identity of the combatants or the nature of the bone of contention, and therefore could not directly test the countervailing effects of diversity on conflict. My contribution highlights that the issue over which combatants fight is a crucial dimension to consider when studying the determinants of conflict.

The analysis at the ethnicity level allows me to delve into the mechanisms through which cultural distance generates conflict over government power. I document that reaction to cultural distance occurs over and above the effect of exclusion from power or discrimination, suggesting that another mechanism is at play. I show that culturally distant groups are more likely to disagree over the mix of public policies, and are thus more likely to rebel to wrest control of the central government.

This paper focuses on ethnic civil conflict and, as a consequence, on a group's cultural distance to the ethnic groups holding government power. Civil conflict, while largely prevalent, is only one type of the many possible manifestations of inter-group violence. A promising avenue for future research is to test the hypothesis of this paper in the context of non-civil conflict, i.e., inter-group violence not involving the government, so far largely overlooked by the literature.[40] Another promising area for future research is improving the measurement of cultural distance. Linguistic distance is by now a well-established measure, but it does not allow to test which exact dimension of culture triggers conflict. Efforts to shed light on this issue are underway,[41] but exploring new metrics that allow to separately examine different cultural components is a promising area for future lines of inquiry.

Taken together, my findings contribute to the literature on diversity and conflict by showing that, contrary to the most pessimistic primordialist views, cultural distance does not always and inevitably come with more conflict. Cultural distance triggers one specific type of dispute—i.e., the one over government power—, and it does so through a specific mechanism—i.e., by generating diverging preferences. As such, my analysis speaks against the widespread idea that ethnic conflict is the sole result of deep-rooted and insurmountable ethnic hatred between groups, and rather suggests that a mix of public policies accommodating diverging preferences might be powerful in alleviating conflict in multicultural societies.

[40] A notable exception is the work by Depetris-Chauvin and Özak [2020].

[41] See, for example, Guarnieri and Tur-Prats [2020], who focus on distance gender norms and investigate its impact on the intensive margin of violence.

Appendices

Appendix A

Appendix to Chapter 1

A.1 Historical Background

By building upon historical colonial records, this section describes in detail the spectrum of colonial policies that characterized the British and French Cameroons in the almost 42 years of colonial partition. The focus lies on how Anglo-French differences in the institutional and legal setting, in the education system and in labor-market policies affected the position of women in society.

A.1.1 Institutional Setting

From the institutional viewpoint, the British administration was characterized by the practice of indirect colonial rule, which allowed the incorporation of native chiefs into the local political system. As highlighted in Fonchingong and Fonjong [2003], this practice encouraged general activism of natives in public life, and boosted development also via an increased active cooperation of women in local initiatives. Female participation in political life, already observed during the 1955 United Nations Visiting Mission (United Nations Trusteeship Council [1956], p.139), was formalized by the Southern Cameroons Electoral Regulations in 1957, which stated that "*[women] may vote and stand as candidates for election under the same conditions as men.*" (Government of the United Kingdom [1958b], par. 635 and 991). In addition to this, women were regarded as men in front of the law and were entitled to acquire and hold their own property, as prescribed in the Married Women's Property Acts introduced in the United Kingdom in 1882. More specifically, the colonial report to the league of Nations reads:

> *"A woman may sue and be sued in the courts as though she were a man, and a married woman is in this respect in the same position as a single woman. The status of single women has never been essentially different from that of men in any branch of the law of property. A married woman is capable of acquiring, holding, and disposing of by will or otherwise and real or personal property as if she were a single woman and any earnings and property acquired by her are her separate property."*—(Government of the United Kingdom [1958b], par. 643)

As it was the case already in their other colonies, the British implemented a legal system based on common law, which is still applied in the Anglophone regions of today's Cameroon and—compared to the civil law system of the French—has been linked to better government

performance, higher public good provision, and more secure property rights (La Porta *et al.* [1999]).

On the other side of the border, the French applied a system of direct rule based on the policy of assimilation, in turn guided by the mission to civilize Africans according to Western principles and institutions (Le Vine [1964], p. 91). This practice led to the gradual abolition of the traditional power of the chiefs, who were substituted with regional figures (*chef de region*) appointed by the French administrators based on their willingness to be subservient to the French authority. The 1922 French colonial report to the League of Nations (cited in Le Vine [1964], p. 95) reads: "*The regional chiefs, a creation of the French administration, have only the authority which is delegated to them; they have no power of their own*". As a consequence, the French areas were characterized by far less vigorous local institutions than the British ones: according to Lee and Schultz's [2012] contribution, this is one of the crucial factors underlying the wealth differential between former British and French areas today.

The French implemented a civil law system, but combined with a practice of legal differentiation, which applied different policies and standards to Africans according to their advancement in the "evolution" towards the French ideal: until 1946, separate legal systems were assigned to *citoyens*, assimilated to European law, and to the so-called *sujets*, subject instead to native customary rules. Only *sujets* upgrading to Gallic standards via education or employment of European character could become *citoyens* (Le Vine [1964], p. 99).According to Le Vine [1964], such practice was actually in place until 1952, when the forced labor system was officially abolished (see Section A.1.3). The evolution to the privileged *citoyens* status was precluded to women, who faced restricted opportunities for formal employment and education. As such, women remained subject to local custom, which, despite allowing women to own their own property, granted them few rights and did not accord them an independent legal status (United Nations Trusteeship Council [1949], United Nations Trusteeship Council [1956]).

The limited regard towards female public and social participation persisted throughout the colonial period, and culminated in a recommendation from the Trustee Council of the League of Nations to the French administration to *"make efforts to improve the social status of women in the Territories*" (United Nations Trusteeship Council [1950a]).

A.1.2 Education

The two colonial powers markedly differed in the education systems they promoted. The British allowed English speaking Protestant missions to monopolize the supply of education. The French, parallel to predominantly Catholic mission schools, instituted a public system aimed merely at the education of a restricted elite of native young men, who would become functional to the colonial administration. In the early years, the primary concern is stressed in the 1923 French colonial report: "*The first official organization of formal education, aimed at meeting the urgent needs of teachers and local administrative staff, has reserved the greatest place for boys' instruction*" (Gouvernement Francais [1923], p.24). As the installation period elapsed, the French authorities acknowledged the necessity to engage in the education of women and established the so-called *écoles ménagères* (housekeeping schools). Education for girls took the restricted form of training within the boundaries of the domestic sphere:

> *"The practical purpose of teaching girls is to teach the pupil what, after marriage, will be her duty as a mother [...]. We want to train housewives [...]. In a word, we propose to create the sense of a new home where the woman holds the place assigned to her by her family and social functions"*—(Gouvernement Francais [1923], pp. 24-25)

This attitude towards female education—expressed by the French administration in the early years of domination—translated into low school attendance of girls throughout the colonial period. The United Nations Mission report after the visit to French Cameroon in 1952 highlights the following:

> *"The Mission [...] observed that the number of girls attending school was about 20 per cent of the total number of pupils in the Territory. The Mission [...] frequently took the occasion to stress the importance to the advancement of the Territory of educating girls. It hopes that the Administration will increase its efforts in this field as far as possible."*—(United Nations Trusteeship Council [1954], par. 269-271)

Despite documented investments in the construction of schools in the four-year development plan between 1953-1956, and consequent increases in overall enrollment rates, the 1958 United Nations Visiting Mission fails to report a remarkable positive impact on girls' education. After independence, Le Vine [1964] stresses how the lag in the education of women—in addition to low completion rates and the inadequacy of most teachers—are to be acknowledged as the limitations of the French education system in former East Cameroon.

A different picture emerges when turning to the Cameroonian regions colonized by the British. The task of educating the natives was assigned to missionary schools based on Protestant principles. Across sub-Saharan Africa, Nunn [2014a] has shown that Protestant missionary activities had a long-lasting impact on the education of girls, as opposed to Catholic schools, which exerted a greater impact on male education in the long run. Protestantism encouraged girls' education, based on Luther's idea that women had to be able to read the Bible in order to go to heaven. Increasing enrollment rates of girls in the British areas of Southern Cameroon are documented as follows:

> *"Enrollment figures indicate that the prejudice against the education of girls is being rapidly overcome. One Mission has reported the significant fact that enrollment of girls in the highest primary class has increased by 66 percent during the past 3 years, which points to the fact that parents are now prepared to keep girls at school for at least the full primary course [...]. One Voluntary Agency reports that in some classes in mixed schools the girls outnumbered the boys."*—(Government of the United Kingdom [1958b], par. 641)

Similarly to the French case, the domestic aspect of education for girls was also present in British schools, but more in the form of a specialization within a unique curriculum:

> "*There are no differences between the education of girls and that of boys, except that the girls often take Domestic Science instead of Rural Science [...]. In nearly all schools boys and girls are taught together in all classes*"—(Government of the United Kingdom [1958b], par. 641)

A.1.3 Employment Opportunities

Finally, let us consider the divergent labor policies in place under the two colonial administrations. After the partition of Cameroon at the end of World War I, the British abolished the German practice of forced labor, and introduced cash wages. Plantation labor opened up unprecedented employment opportunities for the local population, and in particular for women. The British imported their expertise in tea cultivation from their Asian colonies, along with the respective employment customs:

> *"The then expatriate managers were hardly familiar with gender roles and relations in Africa. However, they were often acquainted with, and had been employed on, tea estates in India and Sri Lanka, where plucking was done mainly by women. If women in Assam were plucking tea, why could women in Cameroon not do so? Due to the high employment rate of women on tea estates in Asian countries and elsewhere, tea plucking had to a large extent come to be identified as women's work"*—(Konings [2012])

The increased participation of women in the labor market is documented in the British colonial report of 1958, which highlights the newly gained opportunities for women to earn cash wages under the same conditions as their male counterparts (Government of the United Kingdom [1958b], par. 637). This fact, together with women's ability to hold their earned income separately from their husbands, led the British colonial administration to diagnose a trend towards an increasing independence of women in Southern Cameroon (Government of the United Kingdom [1958b], par. 637).

In order to efficiently manage plantation labor, the British administration established the Cameroon Development Corporation in 1947. During colonial times, the Corporation employed more than half of wage earners in Southern Cameroons (Government of the United Kingdom [1958b], par. 650). It continued to exist also after independence, specializing in the production of tea, palm products, bananas, cocoa and rubber for export purposes. Nowadays, it constitutes the second largest employer after the state of Cameroon. A section on the employment of women in the 1958 report confirms the attractiveness of the new salaried employment opportunities:

> *"For some time the Corporation has been encouraging the employment of women in all grades, and following the higher rates paid to women labor during October, 1957, an increasing number of women have come forward for field duties, some as casual labor, others accepting full-time employment."* —(Government of the United Kingdom [1958a], p. 14)

Finally, the colonial report documents the engagement of women in professional activities also in the public sector and in commercial firms (Government of the United Kingdom [1958b], par. 638).

On the other hand, the French administration kept the extractive labor system introduced by the Germans. Under the form of a *prestation*, it consisted in the obligation to supply ten days of free labor a year for Cameroonian men of *sujet* status. The main goal of the French administration in the first decade of domination was to recruit the necessary workforce to implement large scale infrastructure projects. Workers were recruited by the local administration, and had to endure harsh conditions (Le Vine [1964], pp. 104-110). Despite the exemption of women and children from the *prestation* (Gouvernement Francais [1927], p. 86), instances were reported of coercive use of women for porterage and of children for roadwork (Le Vine [1964], p. 110). The forced labor system did not completely disappear from French Cameroon until 1952. However, according to Le Vine [1964], even after the official abolition of forced labor and of legal differentiation between *cytoyens* and *sujets* at the beginning of the Fifties, *"the old system continued to color relations between the indigenous elites in areas where former extractions had been heaviest"*. As documented by a strand of the economics literature for a variety of contexts, such extractive labor practices are likely to have persistent effects in the long run, not only on labor market-related outcomes, but also on economic development in general (Dell [2010], Lowes and Montero [2020]).

As far as women's participation in the labor market is concerned, the early colonial reports document that "*le travail féminin n'existe pas*" (Gouvernement Francais [1922], p. 25), i.e.

female labor does not exist. Even in the export-oriented agricultural sector, among the salaried labor force, no employment of women was taking place.

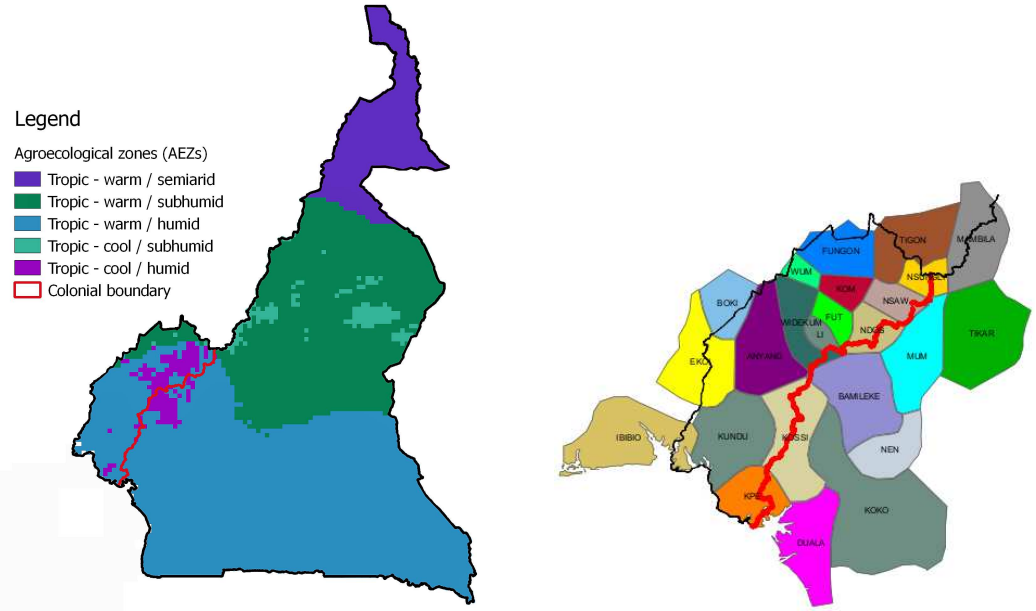
> "*Il est apparu que dans toutes les régions exportatrices de produits où l'on constate un emploi de main-d'oeuvre salariée, l'emploi des femmes pour un travail agricole ou industriel était nul.*"—"*It appeared that in all regions exporting products with a hired labor force there was no employment of women for agricultural or industrial work.*" (Gouvernement Francais [1923], p.14)

For the subsequent decades of the colonization period, historical records do not document any effort of the French administration towards the creation of opportunities for paid employment comparable to those experienced by Cameroonian women on the British side of the boundary. This is consistent with the aforementioned purpose of the education policy of the French administration, i.e. the relegation of women to the domestic sphere.

When tackling the issue of low productivity of indigenous labor in French territories—reported by the United Nations Visiting Mission in 1952—the administration adopted measures aimed at increasing workers' technical abilities, but mainly in male-dominated sectors. Vocational training, instituted by the French colonial government, by Catholic missions and private enterprises, was centered around the activities of masonry, carpentry, joinery and brick work (United Nations Trusteeship Council [1954], par. 227-228). Moreover, during the decade between 1946 and 1957, French Cameroon benefited from large investment flows from the metropole, financed via the FIDES investment fund (*Fonds d'Investissements pour le Développement Économique et Social des territoires d'Outre-Mer*) and structured into two development plans. The main purpose of the latter was to expand and improve the existing infrastructure put into force in the early colonial period. The recruitment of the necessary (male) labor force partly occurred, at least until 1952, under the coercive means described above. The plan materialized with the construction of new roads, bridges and dams, and the improvement of the railway network and port facilities (United Nations Trusteeship Council [1958]).

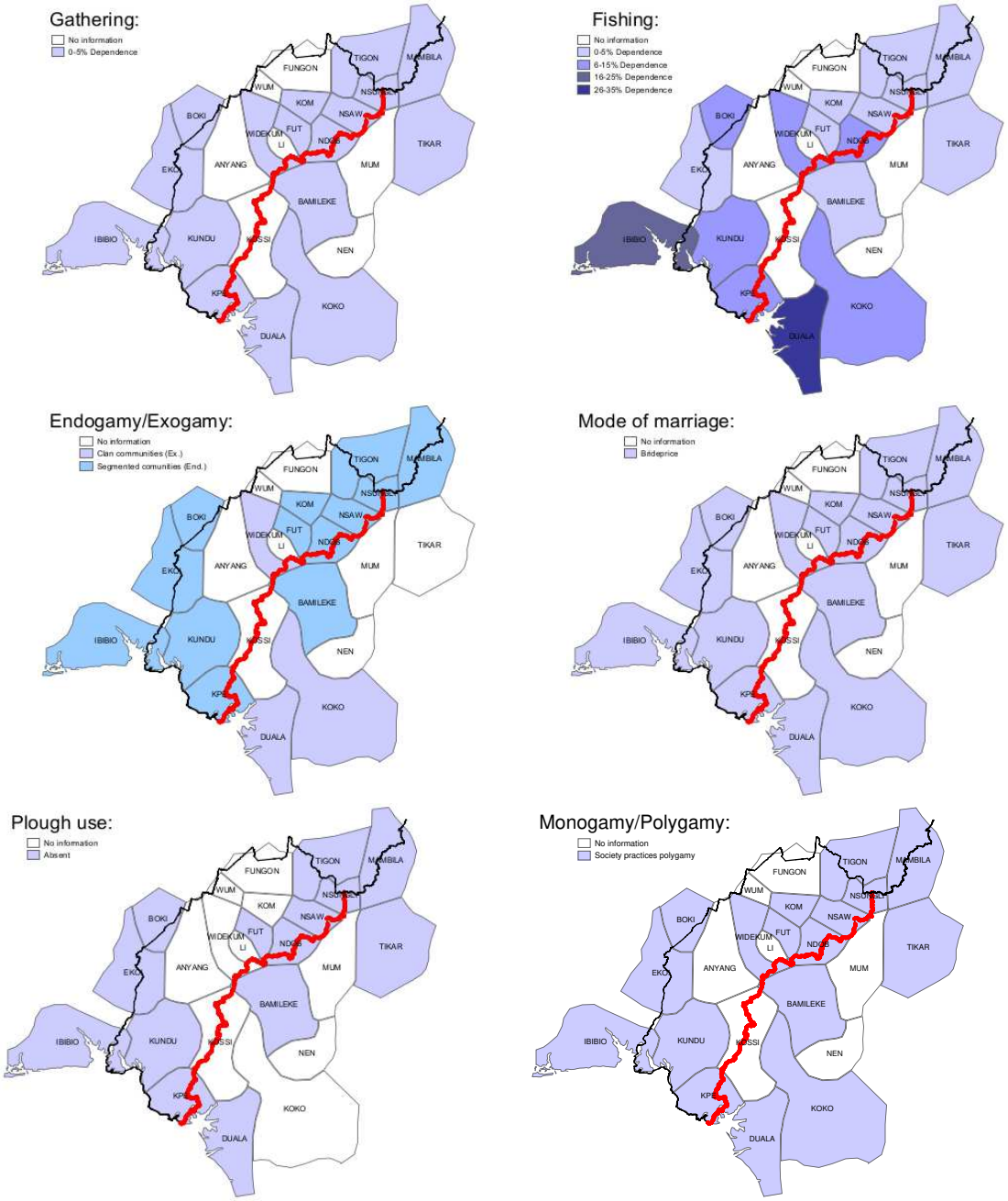
A.2 Additional Figures and Tables

Figure A.1: *Randomness of the Border: Agro-Ecological Zones and Ethnic Settlements*



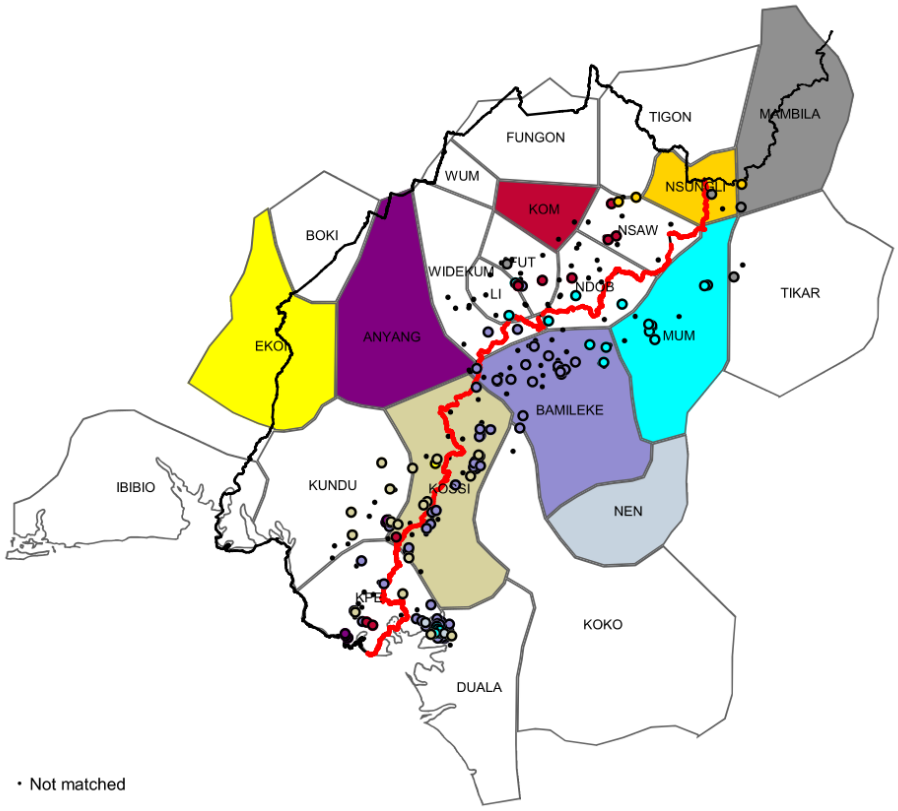
Notes: The map on the left overlays the colonial border to the agro-ecological zones (AEZs) as classified by FAO and developed by the International Food Policy Institute (International Food Policy Research Institute (IFPRI) [2015]) for sub-Saharan Africa. Geographical locations falling under the same AEZ category are characterized by similar climatic characteristics in terms of rainfall and temperature, and therefore provide the same agricultural potential. The map on the right superimposes the colonial boundary with the ethnic pre-colonial settlements in Murdock [1967] using the dataset provided by Nunn and Wantchekon [2011].

Figure A.2: *Pre-Colonial Ethnic Characteristics*



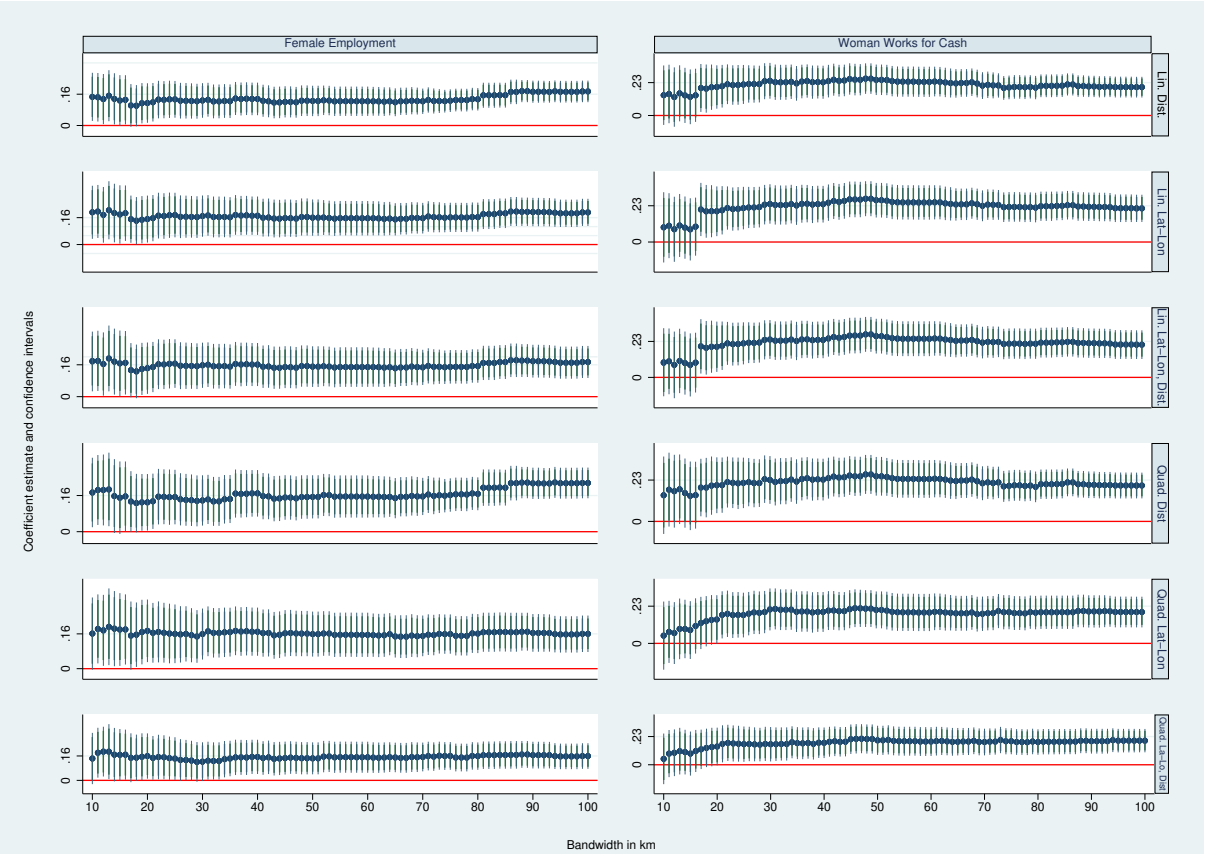
Notes: The maps superimpose the colonial boundary (in red), nowadays borders (in black) with ethnic settlements in Murdock's Ethnographic Atlas (Murdock [1967]), in the version provided by (Nunn and Wantchekon [2011]). Each panel represents pre-colonial characteristics related to: dependence on gathering, dependence on fishing, the practice of endogamy, mode of marriage, the use of the plough and polygamy. Different colors denote different categories for every variable, described in legends specific to each panel. White-filled areas denote lack of information.

Figure A.3: *Contemporaneous and Pre-colonial Settlements*



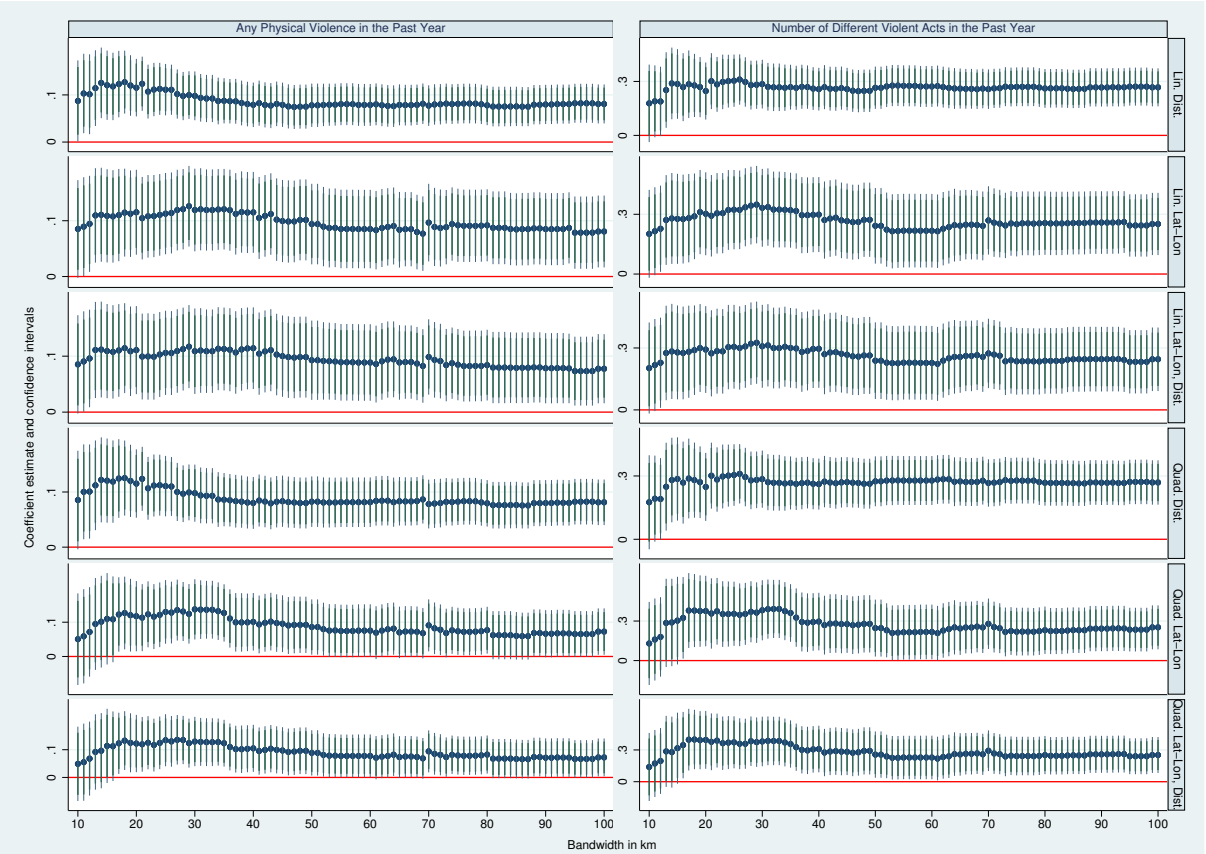
Notes: The map superimposes DHS clusters with the Ethnographic Atlas, assigning to each cluster a different color according to the prevalent ethnicity inhabiting it. 41% of clusters were successfully assigned to a group in the Ethnographic Atlas: not merged clusters are depicted in black.

Figure A.4: *Robustness of Employment Estimates*



Notes: Each sub-graph reports coefficient estimates and confidence intervals of Equation 2.1 (vertical axis) for different bandwidth levels ranging from 10 to 100 kilometers (horizontal axis). Figures on the left column refer to female employment as outcome variable, while figures on the right column refer to women working for cash. Each row corresponds to alternative specifications of the RD polynomial, and alternative choices of the running variable. Thin gray bars represent 95% confidence intervals, while thicker green bars represent 90% confidence intervals, both arising from clustering standard errors at the DHS-cluster level. Calonico *et al.* [2019] optimal bandwidth using distance as running variable ranges between 25 and 40 kilometers depending on the outcome considered and the covariates specified. 40 kilometers best reflects the spatial setting, because it arises from including border and year fixed effects as covariates.

Figure A.5: *Robustness of IPV Estimates*



Notes: Each sub-graph reports coefficient estimates and confidence intervals of Equation 2.1 (vertical axis) for different bandwidth levels ranging from 10 to 100 kilometers (horizontal axis). Figures on the left column refer to the violence dummy outcome variable, while figures on the right column refer to the violence index. Each row corresponds to alternative specifications of the RD polynomial, and alternative choices of the running variable. Thin gray bars represent 95% confidence intervals, while thicker green bars represent 90% confidence intervals, both arising from clustering standard errors at the DHS-cluster level. Calonico *et al.* [2019] optimal bandwidth using distance as running variable ranges between 25 and 40 kilometers depending on the outcome considered and the covariates specified. 40 kilometers best reflects the spatial setting, because it arises from including border and year fixed effects as covariates.

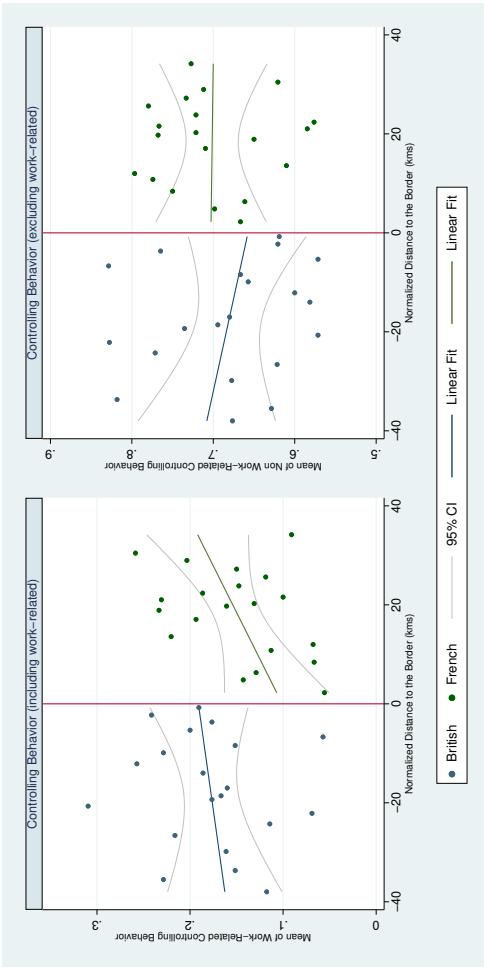


FIGURE A.6: *Controlling Behavior: RD Plots*

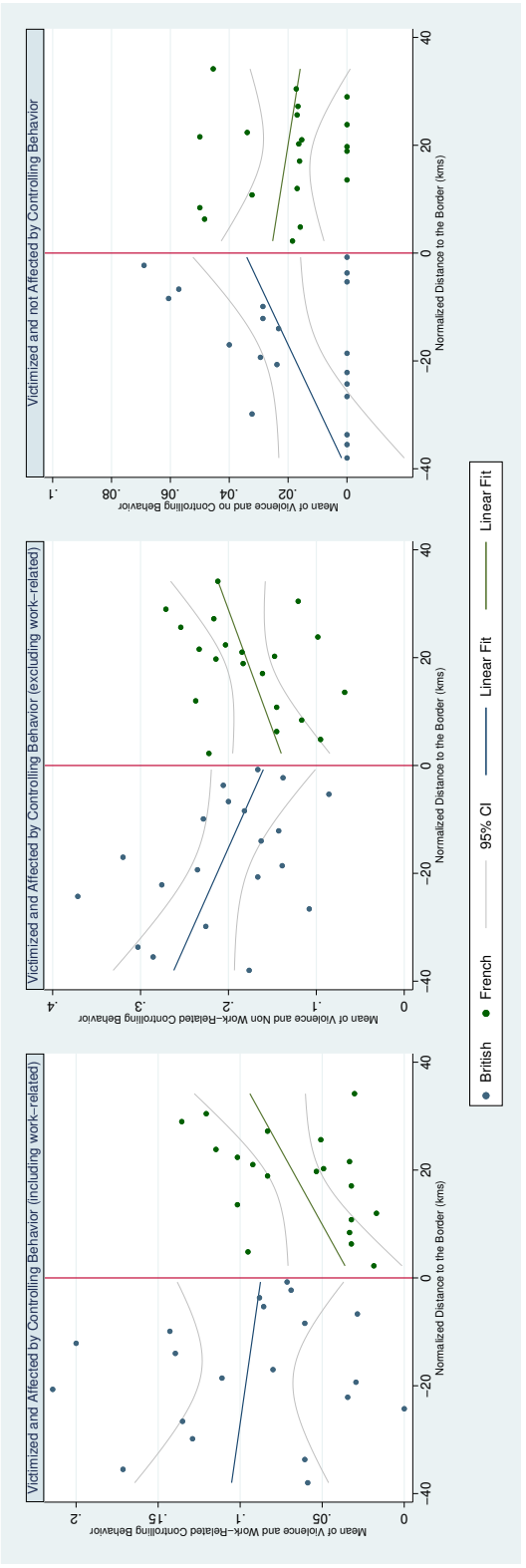


FIGURE A.7: *Victimization and Controlling Behavior: RD Plots*

TABLE A.1: *Robustness of Female Employment Estimates and Placebo Tests*

	Robustness Checks											Placebo Tests			
	Interact. Lin. Dist.	Quadr. Dist.	Cubic Dist.	Quadr. Lat-Lon	Cubic Lat-Lon.	Tri. Kernel	Donut 5 km	No Douala	Add. Contr.	Ethn. FE	WaveSeg FE	Shift West Dist	Shift West Lat-Lon	Shift East Dist	Shift East Lat-Lon
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Panel A: Employment															
British ($\hat{\gamma}$)	0.166***	0.170***	0.153***	0.170***	0.147***	0.156***	0.203***	0.159***	0.148***	0.155***	0.153***	0.039	0.013	0.035	0.058
	(0.054)	(0.052)	(0.052)	(0.056)	(0.057)	(0.054)	(0.065)	(0.056)	(0.055)	(0.055)	(0.052)	(0.046)	(0.042)	(0.061)	(0.070)
Mean	0.730	0.730	0.730	0.730	0.730	0.730	0.719	0.777	0.724	0.729	0.730	0.867	0.867	0.658	0.658
Observations	2,030	2,030	2,030	2,030	2,030	2,097	1,824	1,507	1,936	2,023	2,030	970	970	1,502	1,502
Panel B: Paid Employment															
British ($\hat{\gamma}$)	0.169***	0.232***	0.228***	0.194***	0.129**	0.175***	0.295***	0.225***	0.177***	0.222***	0.227***	-0.163	-0.123	0.069	0.061
	(0.060)	(0.059)	(0.062)	(0.063)	(0.063)	(0.054)	(0.067)	(0.062)	(0.065)	(0.058)	(0.060)	(0.102)	(0.095)	(0.061)	(0.066)
Mean	0.622	0.622	0.622	0.622	0.622	0.622	0.618	0.643	0.625	0.622	0.622	0.705	0.705	0.557	0.557
Observations	2,030	2,030	2,030	2,030	2,030	2,246	1,824	1,507	1,936	2,023	2,030	970	970	1,502	1,502
Year & Bord. FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
YearXBord. FEs											Yes				
Geo ctrls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Baseline ctrls									Yes						
Ethnicity FEs										Yes					

NOTES: OLS coefficient estimates for various robustness and placebo tests. Columns (1)-(5) control for different RD polynomials. Column (6) reports and estimate for Calonico *et al.* [2019] data-driven bandwidth using an interacted linear polynomial in distance and a triangular kernel. Column (7) displays coefficients for a donut specification that excludes observations within 5 km of the boundary. Column (8) excludes observations located in Douala, the economic capital. Column (9) includes controls for altitude, climate, ancestral dependence on fishing, ancestral practice of endogamy, a dummy for urban, and a measure of exposure to slave trade (ln(1+slave exports/land area)). Column (10) controls for ethnicity fixed effects. Column (11) controls for border fixed effects interacted with year fixed effects. Columns (12)-(15) conduct placebo tests and shift the border by 40 kilometers. Robust standard errors, clustered by DHS survey cluster, are reported in parenthesis. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

TABLE A.2: *Robustness of IPV Estimates and Placebo Tests*

	Robustness Checks											Placebo Tests			
	Interact. Lin. Dist.	Quadr. Dist.	Cubic Dist.	Quadr. Lat-Lon	Cubic Lat-Lon.	Tri. Kernel	Donut 5 km	No Douala	Add. Contr.	Ethn. FE	WaveSeg FE	Shift West Dist	Shift West Lat-Lon	Shift East Dist	Shift East Lat-Lon
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Panel A: Any physical violence last year															
British ($\hat{\gamma}$)	0.135***	0.116***	0.105**	0.124***	0.148***	0.113***	0.137**	0.102**	0.182***	0.124***	0.121***	-0.000	-0.004	0.024	0.064
	(0.046)	(0.043)	(0.043)	(0.046)	(0.051)	(0.039)	(0.055)	(0.043)	(0.047)	(0.044)	(0.042)	(0.088)	(0.078)	(0.048)	(0.051)
Mean	0.283	0.283	0.283	0.283	0.283	0.285	0.285	0.281	0.283	0.283	0.283	0.335	0.335	0.266	0.266
Observations	2,030	2,030	2,030	2,030	2,030	2,264	1,824	1,507	1,936	2,023	2,030	971	971	1,502	1,502
Panel B: Number of different violent acts															
British ($\hat{\gamma}$)	0.306***	0.326***	0.313***	0.352***	0.351***	0.264***	0.404***	0.302***	0.440***	0.324***	0.337***	0.054	-0.006	0.061	0.091
	(0.114)	(0.108)	(0.110)	(0.116)	(0.125)	(0.094)	(0.143)	(0.111)	(0.119)	(0.111)	(0.110)	(0.261)	(0.226)	(0.107)	(0.119)
Mean	0.576	0.576	0.576	0.576	0.576	0.580	0.591	0.579	0.581	0.577	0.576	0.745	0.745	0.507	0.507
Observations	2,030	2,030	2,030	2,030	2,030	2,402	1,824	1,507	1,936	2,023	2,030	971	971	1,502	1,502
Year & Bord. FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
YearXBord. FEs											Yes				
Geo ctrls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Baseline ctrls									Yes						
Ethnicity FEs										Yes					

NOTES: OLS coefficient estimates for various robustness and placebo tests. Columns (1)-(5) control for different RD polynomials. Column (6) reports and estimate for Calonico *et al.* [2019] data-driven bandwidth using an interacted linear polynomial in distance and a triangular kernel. Column (7) displays coefficients for a donut specification that excludes observations within 5 km of the boundary. Column (8) excludes observations located in Douala, the economic capital. Column (9) controls for altitude, climate, ancestral dependence on fishing, practice of endogamy, urban, and exposure to slave trade (ln(1+slave exports/land area)). Column (10) controls for ethnicity fixed effects. Column (11) controls for border fixed effects interacted with year fixed effects. Columns (12)-(15) conduct placebo tests and shift the border by 40 kilometers. Standard errors clustered by DHS cluster are reported in parenthesis. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

Table A.3: *Linking the Discontinuities for IPV, Coercive Partner Behavior, and Paid Employment*

	$P(V \cap C \cap W)$			$P(V \cap C \cap W \cap E)$			$P(V \cap C \cap W \cap \not{E})$		
	Dist	Lat-Lon	Dist Lat-Lon	Dist	Lat-Lon	Dist Lat-Lon	Dist	Lat-Lon	Dist Lat-Lon
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
British ($\hat{\gamma}$)	0.074***	0.076***	0.077***	0.053***	0.055**	0.056***	0.021	0.021	0.021
	(0.023)	(0.025)	(0.025)	(0.020)	(0.022)	(0.021)	(0.013)	(0.014)	(0.014)
Mean	0.080	0.080	0.080	0.056	0.056	0.056	0.023	0.023	0.023
Obs	1,897	1,897	1,897	1,897	1,897	1,897	1,897	1,897	1,897
Clusters	357	357	357	357	357	357	357	357	357

Notes: OLS estimates of Equation (2.1) for a window of 40 kilometers around the boundary. Standard errors clustered by DHS survey cluster are reported in parentheses. All regressions include geographic controls (distance to the capital and distance to Douala), together with border segment fixed effects. In columns (1)-(3) we report estimates from Table 1.5, i.e. the border effect on the joint probability of being victimized and subject to work-related coercive behavior, $P(V \cap C \cap W)$. Columns (4)-(9) further decompose this discontinuity into two parts: the joint probability of being victimized, subject to work-related coercive behavior, and employed, $P(V \cap C \cap W \cap E)$, and the joint probability of being victimized, subject to work-related coercive behavior, and not employed, $P(V \cap C \cap W \cap \not{E})$. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

Table A.4: *Other Outcomes Consistent with Women's Economic Empowerment*

	A. Decision over Purchases			B. Own Property		
	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon
	(1)	(2)	(3)	(4)	(5)	(6)
British ($\hat{\gamma}$)	0.148***	0.185***	0.185***	0.068**	0.070**	0.074**
	(0.053)	(0.057)	(0.055)	(0.032)	(0.030)	(0.031)
Mean	0.710	0.710	0.710	0.080	0.080	0.080
Observations	1,880	1,880	1,880	1,592	1,592	1,592
Clusters	358	358	358	196	196	196

Notes: OLS estimates of Equation (2.1) for a sample of women within 40 kilometers from the boundary. *Decision over Purchases*: 1 if women report having a say about household decisions over purchases. *Own Property*: 1 if a woman owns land or a house with a title (variable available only for the 2011 DHS wave). Each column controls for a different specification of the RD polynomial. All regressions include geographic controls (distance to the capital and distance to Douala), together with border segment fixed effects. Columns (1)-(3) includes year fixed effects. Standard errors clustered by DHS survey cluster are reported in parentheses. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

Table A.5: *Household Wealth*

	Dependent Variable: Household Wealth Index											
	A. Women in Domestic Violence Module			B. All Women			C. All Men			D. All Women and Men		
	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist Lat-Lon	Dist
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
British ($\hat{\gamma}$)	0.428**	0.365*	0.434**	0.529***	0.500***	0.535***	0.416*	0.349**	0.437**	0.512***	0.453***	0.521***
	(0.215)	(0.188)	(0.205)	(0.197)	(0.174)	(0.189)	(0.212)	(0.176)	(0.198)	(0.194)	(0.172)	(0.184)
Mean	3.491	3.491	3.491	3.627	3.627	3.627	3.737	3.737	3.737	3.662	3.662	3.662
Observations	2,030	2,030	2,030	9,044	9,044	9,044	4,209	4,209	4,209	13,253	13,253	13,253
Clusters	358	358	358	361	361	361	361	361	361	361	361	361

Notes: OLS estimates of Equation (2.1) for a sample of clusters within 40 kilometers from the boundary. The outcome variable is the DHS household wealth index (ranging between 1 and 5). Columns (1)-(3) include the sample of women interviewed in the domestic violence module; columns (4)-(6) include all women; columns (7)-(9) include all men interviewed in the DHS men's module; columns (7)-(9) include all women and all men. Each column controls for a different specification of the RD polynomial. All regressions include geographic controls (distance to the capital and distance to Douala), together with border segment fixed effects and year fixed effects. Standard errors clustered by DHS survey cluster are reported in parentheses. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

Table A.6: *Informal Unions, Relationship Duration, and Separation*

	A. Informal Union			B. Marital Duration			C. Informal Separation			D. Divorce		
	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist Lat-Lon	Dist
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
British ($\hat{\gamma}$)	-0.050	-0.038	-0.050	-0.059	0.006	-0.059	-0.001	-0.016	-0.002	0.003	0.003	0.003
	(0.034)	(0.033)	(0.034)	(0.162)	(0.158)	(0.162)	(0.021)	(0.020)	(0.021)	(0.010)	(0.010)	(0.010)
Mean	0.190	0.190	0.190	2.973	2.973	2.973	0.0527	0.0527	0.0527	0.0143	0.0143	0.0143
Observations	2,030	2,030	2,030	2,030	2,030	2,030	2,030	2,030	2,030	2,030	2,030	2,030
Clusters	358	358	358	358	358	358	358	358	358	358	358	358

Notes: OLS estimates of Equation (2.1) for a sample of women within 40 kilometers from the boundary. *Informal Union*: 1 if a woman is living together with her partner, 0 otherwise. *Marital Duration*: variable ranging between 1 and 7 capturing the duration of a relationship with 1 denoting 0-4 years, 2 denoting 5-9 years, 3 denoting 10-14 years, 4 denoting 15-19 years, 5 denoting 20-24 years, 6 denoting 25-29 years,7 denoting 30+ years. *Informal Separation*: 1 if a woman is informally separated, i.e., if she does not live together with her partner, 0 otherwise. *Divorce*: 1 if a woman is divorced, 0 otherwise. Each column controls for a different specification of the RD polynomial. All regressions include geographic controls (distance to the capital and distance to Douala), together with border segment fixed effects and year fixed effects. Standard errors clustered by DHS survey cluster are reported in parentheses. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

Table A.7: *Women's Sectoral Employment*

	A. Women Agriculture			B. Women Manual			C. Women Services		
	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
British ($\hat{\gamma}$)	0.096	0.106	0.094	0.023	0.019	0.024	0.019	0.016	0.019
	(0.077)	(0.070)	(0.074)	(0.039)	(0.038)	(0.039)	(0.018)	(0.017)	(0.018)
Mean	0.333	0.333	0.333	0.206	0.206	0.206	0.031	0.031	0.031
Observations	2,026	2,026	2,026	2,026	2,026	2,026	2,026	2,026	2,026
Clusters	358	358	358	358	358	358	358	358	358

Notes: OLS estimates of Equation (2.1) for a sample of women within 40 kilometers from the boundary. *Women Agriculture*: 1 if woman is employed in the agricultural sector, 0 otherwise; *Women Manual*: 1 if woman is employed in a manual occupation, either skilled or unskilled, 0 otherwise; *Women Services*: 1 if woman is employed in services, 0 otherwise. Each column controls for a different specification of the RD polynomial. All regressions include geographic controls (distance to the capital and distance to Douala), together with border segment fixed effects and year fixed effects. Standard errors clustered by DHS survey cluster are reported in parentheses. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

Table A.8: *Other Measures of Women's Educational Attainment*

	A. Completed Primary			B. Completed Secondary			C. Literacy		
	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon	Lat-Lon	Dist	Dist Lat-Lon
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
British ($\hat{\gamma}$)	0.133**	0.103**	0.134***	0.030	0.021	0.030	-0.086*	-0.074	-0.085*
	(0.052)	(0.051)	(0.051)	(0.025)	(0.023)	(0.024)	(0.049)	(0.052)	(0.047)
Mean	0.779	0.779	0.779	0.067	0.067	0.067	0.817	0.817	0.817
Observations	2,030	2,030	2,030	2,030	2,030	2,030	2,023	2,023	2,023
Clusters	358	358	358	358	358	358	358	358	358

Notes: OLS estimates of Equation (2.1) for a sample of women within 40 kilometers from the boundary. *Completed Primary*: 1 if a woman has completed primary education, 0 otherwise. *Completed Secondary:* 1 if a woman has completed secondary education, 0 otherwise. *Literacy*: 1 if a woman can read a whole sentence or part of it, 0 otherwise. Each column controls for a different specification of the RD polynomial. All regressions include geographic controls (distance to the capital and distance to Douala), together with border segment fixed effects and year fixed effects. Standard errors clustered by DHS survey cluster are reported in parentheses. *** (**) (*) indicates significance at the 1% (5%) (10%) level.

Appendix B

Appendix to Chapter 2

B.1 Data Sources and Dataset Construction

We hereby present in detail the data sources used for the analysis and the procedure adopted to merge the various datasets.

B.1.1 Sexual Violence in Armed Conflict

The source of our dependent variable is the Sexual Violence in Armed Conflict Dataset (SVAC) (Cohen and Nordås [2014]). The SVAC dataset includes information on all conflicts between 1989 and 2009, as defined by the UCDP/PRIO Armed Conflict Database: any "contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths" (Gleditsch *et al.* [2002]). The SVAC dataset provides information on war-related sexual violence perpetrated by three types of armed-conflict actors: government/state military, pro-government militias, and rebel/insurgent forces. In total, the dataset covers 129 active conflicts and 625 armed actors involved in them. Adhering to the International Criminal Court's rationale, SVAC defines war-related sexual violence as including the following acts: rape, sexual slavery, forced prostitution, forced pregnancy, forced sterilization, and forced abortion (International Criminal Court [2000]). In addition, following Wood [2009], sexual mutilation and sexual torture are also included.

SVAC draws upon annual reports from three sources (Amnesty International, Human Rights Watch, and the US State Department) to construct a measure of prevalence of sexual violence at the conflict-actor-year level. The resulting variable is an index ranging between 0-3 that reflects the magnitude of the phenomenon. More specifically, it takes the value 3 if, in a given year of conflict, an actor perpetrated acts of massive, innumerable, or systematic sexual violence according to the aforementioned sources and, furthermore, if reported incidents or victims of sexual violence exceeded 1,000; 2 if sexual violence was described as widespread and common, and reports of victims or incidents ranged between 25 and 999; 1 if reported victims and incidents were below 25 and the occurrence of sexual violence was only isolated; 0 if no sexual violence was mentioned in a given year in relation to a specific conflict.[1]

We exploit the dyadic version of the SVAC dataset, i.e., the GEO-SVAC dataset (Bahgat *et al.* [2016]). GEO-SVAC uses as its starting point the UCDP GED dataset (Sundberg and Melander [2013]; Croicu and Sundberg [2017]) and enriches it with the information on sexual

[1] For further details on the methodology of data collection and coding refer to Cohen and Nordås [2014].

violence provided by SVAC for *state-based* conflicts between government/state military and rebel/insurgent forces between 1989 and 2009.[2]

The unit of observation in GEO-SVAC is a geo-located state-based conflict event. Since the variation in our variable of interest (i.e., sexual violence prevalence) occurs at the actor-conflict-year level—and not at a geo-located event level—we maintain actor-conflict-year as the unit of observation in the analysis. For our purposes, however, GEO-SVAC offers an important advantage. In addition to the identity of the perpetrator of sexual violence, it codes the identity of the other actor involved in the conflict. In other words, the dataset is *dyadic*, i.e. it includes the identity of side A, which is always a government (corresponding "government/state military" in the SVAC coding) and side B, a rebel or opposing government (corresponding to "rebel/insurgent forces" in SVAC). It furthermore reports the intensity of sexual violence perpetrated by both side A and side B in a specific year of conflict.

As illustrated in Figure B.2, we restrict the GEO-SVAC sample in two ways. First, we focus on 45 conflicts fought in the African continent.[3] Second, we restrict our analysis to 33 *ethnic* civil conflicts defined as "armed conflicts between the government of a state and one or more internal opposition group(s) that cause at least 25 battle-related deaths within a year and in which armed groups (i) explicitly pursue ethno-nationalist aims, motivations, and interests; and (ii) recruit fighters and forge alliances on the basis of ethnic affiliation" (Gleditsch *et al.* [2002], Cederman *et al.* [2012]).[4] We categorize conflict-years as ethnic relevant based on Wimmer *et al.*'s [2009] definition.[5] In addition, we include in the sample three additional conflicts that, according to the sources we consulted, qualify as ethnic.[6] Our results are robust to dropping the latter three conflicts from the sample.

B.1.2 Conflict Actors' Ethnic Identity

Next, we assign to each actor (i.e., to both side A and side B in GEO SVAC) an ethnic identity. To achieve this, we exploit the rich information provided by the Ethnic Power Relations (EPR) Dataset Family (Vogt *et al.* [2015]), where an ethnic group is "an identity group that defines itself or is defined by others along linguistic, religious, or racial characteristics". The EPR dataset family provides information, *inter alia*, on ethnic groups' involvement in civil war as part of a rebel organization. We are therefore able to assign to the majority of sides B (i.e. rebel

[2] Figure B.1 illustrates the relationship between GEO-SVAC, UCDP GED, and SVAC. Conflicts covered by GEO-SVAC are only a subset of those included in the original UCDP GED dataset, which includes also *non-state* conflicts and episodes of *one-sided* violence. Moreover, GEO-SVAC does not cover SVAC conflict events involving pro-government militias. As a result, GEO-SVAC includes information on 106 *state-based* conflicts around the world involving the following actors: *government/state military* and *rebel/insurgent forces*. Finally, GEO-SVAC includes only active years of conflict, whereas SVAC provides information also on interim and post-conflict years.

[3] African conflicts constitute 42% of conflicts in GEO-SVAC, which includes a total of 106 conflicts.

[4] "[...] we conducted new research and coded each conflict for whether rebel organizations pursued ethnonationalist aims and recruited along ethnic lines. We also coded whether rebels aimed at establishing a new independent state. We distinguish between ethnic and nonethnic conflicts using the aims of the armed organization and their recruitment and alliance structures [...]. We identify as "ethnic" the aims of achieving ethnonational self-determination, a more favorable ethnic balance-of-power in government, ethnoregional autonomy, the end of ethnic and racial discrimination, language and other cultural rights, and so forth. In ethnic wars, armed organizations also recruit fighters predominantly among their leaders' ethnic group and forge alliances on the basis of ethnic similarity" (Wimmer *et al.*'s [2009]).

[5] In the case of 17 conflicts categorized as ethnic in Wimmer *et al.* [2009], we include additional conflict-years that were not recorded by Wimmer *et al.* [2009] but that were part of a conflict that was qualified as ethnic. Results are robust to excluding these conflict years.

[6] Government of Guinea-Bissau vs. Military Junta for the Consolidation of Democracy, Peace and Justice (1998-1999); Government of Eritrea vs. Government of Ethiopia (1998-2000); Government of Eritrea vs. EIJM-AS (1993-2003).

Figure B.1: *Relationship between UCDP-GED, SVAC and GEO-SVAC*

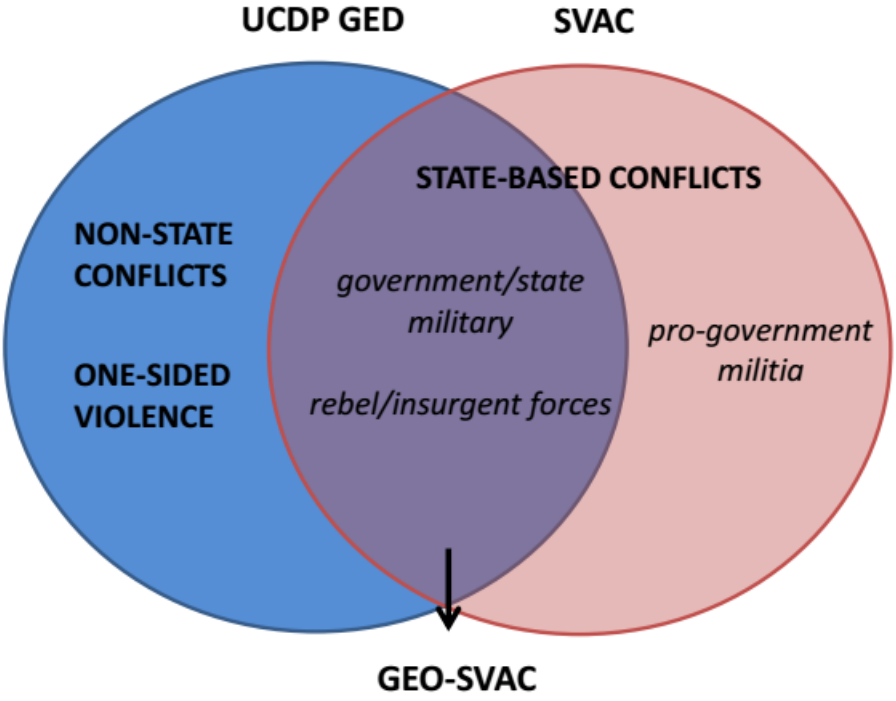
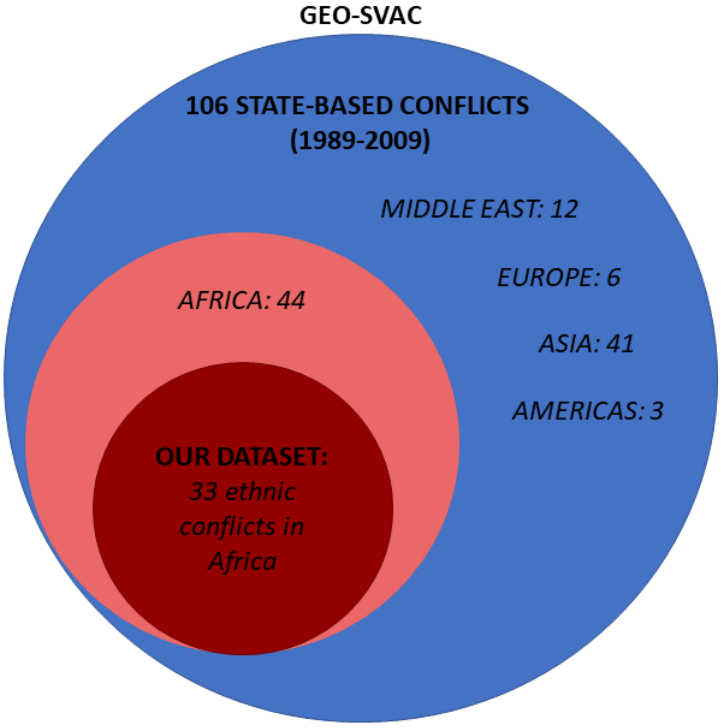


Figure B.2: *Relationship between GEO-SVAC and our Dataset*



or insurgent forces as mentioned above) one or more EPR ethnic groups.[7] Using a variety of additional sources, we identify the ethnic identity of four remaining sides B involved in conflicts classified as ethnic.[8]

The EPR dataset family also provides information on ethnic groups' access to executive government power. When an ethnic group holds exclusive or almost exclusive power in the government, it is classified as either *Monopolist* or *Dominant*. When power is formally or informally shared by different groups, the latter are defined as *Senior* or *Junior Partners*, depending on their relative position in the government. Groups that do not detain any power are either defined as *Discriminated* or *Powerless*, depending on whether or not the central power pursues actions of active discrimination against them. The remaining categories refer to either *Self-Excluded*—i.e. controlling a particular territory in the state that they have declared independent—or *Irrelevant* groups. Finally, EPR also records instances of *State Collapse*.

Given the nature of conflicts included in our data—civil (state-based) ones where side A is always a government—we can assign to side A an ethnic identity. In instances where, in a certain year and during a conflict, central power is held exclusively by one ethnic group—defined by EPR either as *Dominant* or *Monopolist*—the matching is straightforward. Whenever more groups detain government power jointly, we always assign to side A the ethnic identity of the *Senior Partner*, and, in addition, of the *Junior Partner* only in cases where sources indicate direct involvement of that ethnic group in civil conflict. As a result, side A can be assigned to either one or more EPR ethnic groups.[9]

B.1.3 Conflict Actors' Ancestral Characteristics

Finally, we merge our dataset with the Ethnographic Atlas (EA), coded by Murdock [1967] and updated by Nunn and Wantchekon [2011]. The EA is arguably the most compelling source of ethnographic information for 1,265 societies around the world, collected at the end of the 19th century. For Africa, the EA provides detailed information on groups' socio-economic conditions, settlement patterns, and family arrangements prior to European contact.

We link the information provided by the EA to the dataset on conflict through the concordance data provided by Michalopoulos and Papaioannou [2016]. This data links 196 EPR groups to 593 ethnicities in Murdock using a variety of sources. We successfully merge 71 EPR groups to the EA through Michalopoulos and Papaioannou's [2016] concordance table. For 13 of the 15 EPR groups[10] that remain unmerged, we rely on a variety of sources and identify the Murdock groups of interest.[11] For two EPR ethnic groups (Americo-Liberians and Muslim

[7] The sub-dataset—belonging to the EPR dataset family—that allows this merging is ACD2EPR (Wucherpfennig *et al.* [2012]). We classified the ethnic identity of 91 rebel groups with this procedure. To quality-check the validity of the merging, we simultaneously consulted the narratives in the EPR Atlas, accessible through the GrowUp database (https://growup.ethz.ch/). For 51 of these rebel groups, we confirmed the merging by consulting additional sources.

[8] These sides B belong to the three conflicts that Wimmer *et al.* [2009] do not classify as ethnic-relevant (the Military Junta for the Consolidation of Democracy, Peace, and Justice; EIJM-AS; Government of Ethiopia) and to a rebel group whose ethnic identity is missing in ACD2EPR (the AQIM in Algeria and Niger).

[9] We always conduct a quality-check on these merges by consulting the narratives in the EPR Atlas, accessible through the GrowUp database (https://growup.ethz.ch/) and, in the case of 13 (out of 28) governments, additional sources.

[10] These EPR groups are: Afar, Americo-Liberians, Arabs, Arabs/Moors, Bembe, Christian Eritreans, Gio, Goula Isaas (Somali), Mandingo, Masalit, Muslim Eritreans, Somali, Sharawis, and Zaghawa

[11] Sources include the Joshua Project, the Ethnologue dataset, Wikipedia, and others. In some instances, we also exploit the fact that EPR provides the geo-location of the ethnic settlements to cross-validate the just mentioned sources.

Eritreans), it is impossible to identify a correspondence in the EA, and therefore they remain un-merged.

In some cases, this matching procedure results in a one-to-one mapping between EPR and the Ethnographic Atlas. For example, the ethnic group of the rebel force FLEC-FAC in Angola, the *Cabindan Mayombe*, is matched with the *Yombe* group in the EA. However, in other cases, a conflict actor is associated to multiple Murdock groups either because (i) side A, side B, or both are represented by multiple EPR groups, as described in the previous section, or (ii) an EPR group corresponds to multiple groups in the EA, or (iii) both. An example of the latter case is the following: the RFDG rebel group in Guinea is composed of members belonging to the EPR groups called *Malinke* and *Peul*. In turn, the Michalopoulos and Papaioannou's [2016] correspondence table matches *Malinke* to four Murdock groups (*Yalunka, Konyanke, Malinke,* and *Koranko*), and the *Peul* to three Murdock groups (*Foutadjalon, Sokoto, Liptako*). In these instances, we weight the ethnic characteristics of each EPR group by the size of the EA groups to which it corresponds. In the just-mentioned example, *Peul*'s dependence on pastoralism will be a weighted average between *Foutadjalon's, Sokoto's*, and *Liptako's* dependence on pastoralism, based on the three ethnic groups' size, proxied by the land area covered by their settlements. We provide estimates using both the weighted and the un-weighted version of the various ethnic characteristics, and show that our results are generally insensitive to this procedure.

B.1.4 Example

Figure B.3 summarizes the merging process described in this section for a conflict event that took place in 1989 in Liberia between the rebel group NPFL (National Patriotic Front of Liberia) and the government. The GrowUp platform[12] illustrated in Figure B.3 provides a summary of the ethnic power relations in Liberia in the year 1989. One ethnic group, the *Krahn (Guere)*, detains exclusive power in the government and is thus defined as *Dominant*. The remaining politically relevant groups (the *Americo-Liberians*, the *Gio* and the *Mano*) are all discriminated against. However, only the latter two are involved in a conflict, i.e., those marked by a star.[13] Consequently, the group *Krahn (Guere)* is assigned to side A (the government of Liberia), while side B corresponds to the *Gio* and *Mano* groups.

To confirm the validity of these matches, we consult the chapter on Liberia in the EPR Atlas (Girardin *et al.* [2015]). The following extract confirms the *Krahn* dominant position in the government:

> [...] Doe's coup brought an end to the Americo-Liberian dominance. [...] Doe's rule relied heavily on his own *Krahn* group, which occupied the state's key positions. They soon dominated political and military life in Liberia. Thus, the Krahn are coded as "dominant" during Doe's regime. There is also widespread discrimination and state violence against the *Gio* and *Mano* ethnic groups (where opposition against Doe was widespread) [...]. Thus, these groups are also coded as "discriminated".

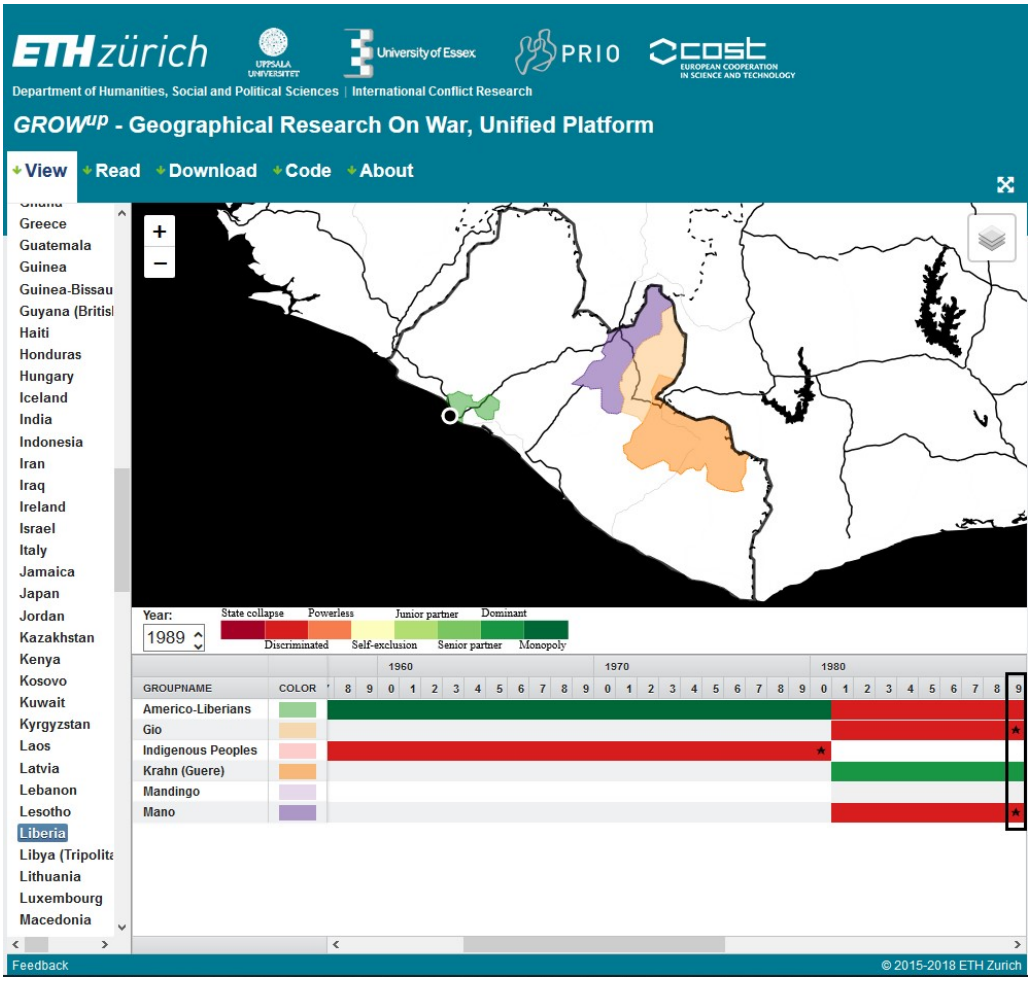
Moreover, to confirm the ethnic nature of the NPFL rebel group, we rely on other narratives, such as Wikipedia, according to which "most NPFL fighters were originally drawn from the Gio and Mano ethnic groups of northern Liberia that were persecuted under Doe's regime".[14]

[12] The interface depicted in Figure B.3 displays the information contained in the Ethnic Power Relations Dataset (Vogt *et al.* [2015]) and the UCDP conflict data (Croicu and Sundberg [2017]).

[13] This is the equivalent of the information contained in the ACD2EPR sub-dataset of the EPR family and in the Wimmer *et al.*'s [2009] classification of ethnic relevant conflict.

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

Figure B.3: *Merging Process Example through the GrowUp Platform*



Notes: The figure displays the Grow-Up platform (Girardin *et al.* [2015]) with the Liberian example. On top, it displays the EPR groups' settlements, and at the bottom the power relations between the six ethnic groups from the 1960s to the 1980s. The year highlighted in black captures the war between the rebel forces represented by the Gio and Mano ethnicities and the government, represented by the Krahn (Guere) ethnic group.

Finally, the remaining step consists in associating the EPR groups with the EA. Straightforwardly, Michalopoulos and Papaioannou [2016] assign *Krahn* to *Kran* and *Mano* to *Ngere* in the Atlas. Gio is not included in the correspondence table. However, we retrieve the necessary information from Holsoe and Lauer [1976], according to whom "in Liberia, Gio persisted as the name for the Dan"[15], and link the EPR group *Gio* to *Dan* in the EA. As a final check, we also compare the EPR ethnic boundaries with the Ethnographic Atlas settlement map.

B.1.5 Linguistic Distance

We use Fearon's [2003] measure of linguistic distance, which is based on linguistic trees in the Ethnologue. For each language, the Ethnologue provides a classification starting with the language family (e.g. Afro Asiatic, Nilo-Saharan, Creole), followed by "nodes", i.e., the branching points of the linguistic tree, and ending with the language itself. We merge information on languages spoken by ethnic groups through the Ethnic Power Relations-Ethnic Dimensions

[15] Page 142 in Holsoe and Lauer [1976].

(EPR-ED) dataset, and compute distances between each pair of languages based on the number of common nodes in the tree.

For example, the language spoken by the ethnic group Zaghawa is classified as follows: *Nilo-Saharan, Saharan, Eastern.* The language of Zaghawa's opponent, Sara, is classified as: *Nilo-Saharan, Satellite-Core, Satellites, Central Sudanic, West, Bongo-Bagirmi, Sara-Bagirmi, Sara, Sara Proper.* These two languages have only one node in common (Nilo-Saharan, i.e. the language family). Following Putterman and Weil [2010], we calculate the distance between language i and language j as follows:

$$d_{ij} = 1 - \left(\frac{\#\text{ of common nodes between } i \text{ and } j}{\frac{1}{2}(\#\text{ of nodes of language } i + \#\text{ of nodes of language } j)} \right)^{\lambda} \tag{B.1}$$

Languages originating from different families have no nodes in common, and their distance will be equal to 1. The parameter λ ranges between 0 and 1, and is used to attribute higher weight to earlier common nodes, as early separations in the language tree are likely to signify larger cultural divergence on average than later separations (see Fearon [2003]). As in Putterman and Weil [2010] and Fearon [2003], we assign to λ the value of 0.5.[16]

The EPR-ED dataset assigns to each EPR ethnic group up to three languages, which are the three largest language segments spoken by group members in descending order. It also attributes a relative size to each of these languages, which sums up to 1 and reflects the percentage of individuals within an ethnic group speaking a specific language. Given this, we exploit the relative size of languages as weights, and calculate the linguistic distance for each perpetrator-victim pair of ethnic groups as follows:

$$LD_{pv} = \sum_{i=1}^{3} \sum_{j=1}^{3} (s_{pi} \times s_{vj} \times d_{ij}) \tag{B.3}$$

where p and v denote an ethnic group on the perpetrator's and on the victim's side, respectively, s_{pi} and s_{vj} denote the relative size of language i (j) in the ethnic group of the perpetrator (victim), and d_{ij} is the linguistic distance between language i and language j described above.

Since perpetrators and victims can be composed by multiple ethnic groups, the ultimate linguistic distance between two opposing actors in a conflict is given by the average distance between each perpetrator-victim ethnic-group pair:

$$LD_{PV} = \sum_{p=1}^{M} \sum_{v=1}^{N} (\frac{1}{M} \times \frac{1}{N} \times LD_{pv}) \tag{B.4}$$

where M denotes the number of ethnic groups fighting on the perpetrator's side, N the number of ethnic groups fighting on the victim's side, and LD_{pv} the linguistic distance of each ethnic group pair.

[16] In the above example, the linguistic distance between Zagawa and Sara is equal to:

$$d_{ij} = 1 - \left(\frac{1}{\frac{1}{2}(3+9)} \right)^{0.5} = 0.59 \tag{B.2}$$

B.1.6 Religious Distance

We construct a measure of religious distance between ethnic belligerents exploiting information on ethnic groups' religion provided by the EPR-ED dataset. Similar to languages, EPR-ED codes up to three religions professed by each ethnic group, as well as their relative size (reflecting the percentage of individuals within an ethnic group professing a specific religion).

We construct a measure of religious distance analogous to the one for linguistic distance (see equations 5-8 in section B.1.5). To this end, we exploit EPR-ED classification of language segments. To continue the example of section B.1.5, the main religion of the ethnic group Zaghawa is Sunni Islam, classified as follows: *Abrahamic Religions, Islam, Sunni Islam.* The main religion of Zaghawa's opponent, Sara, is Protestantism, classified as *Abrahamic Religions, Christianity, Protestantism.* In this case, the two religions in the pair have one node in common, and their distance will be equal to 0.42.[17]

[17]Resulting from equation 5:

$$d_{ij} = 1 - \left(\frac{1}{\frac{1}{2}(3+3)}\right)^{0.5} = 0.42 \tag{B.5}$$

B.2 The Slave Trade and Conflict-Related Sexual Violence

In this section, we test whether an ethnic group's exposure to the transatlantic and Indian Ocean slave trade is a factor explaining the use of sexual violence in conflict, despite not including this ethnic characteristic in the eGII. According to Teso [2018], exposure to a demographic shock such as the transatlantic slave trade, where slaves exported were predominantly men, contributed to the evolution of more gender-equitable norms. In heavily raided ethnic groups, women started taking up typically men's tasks, and this resulted in a shift of the traditional gender division of labor. This shock had persistent effects in the long run: today, women whose ancestors were exposed to the transatlantic slave trade are more likely to be in the labor force, and to have lower fertility and higher decision-making power within the household.

We rely on information on the number of slave shipments provided by Nunn and Wantchekon [2011] to construct a measure of exposure to the transatlantic and the Indian Ocean slave trade. As shown in Table B.1 below, ethnic groups exposed to the transatlantic slave trade are less likely to engage in sexual violence in conflict. Moving from no slave trade exposure to the highest level of exposure in the sample decreases the intensity of sexual violence by 0.86-1.56. Equivalently, one standard deviation increase in exposure to the slave trade reduces sexual violence by 0.09-0.16 standard deviations.

Ethnic groups exposed to the Indian Ocean slave trade, conversely, use sexual violence in conflict at a higher intensity. The Indian Ocean slave trade did not distort the sex ratio as the transatlantic trade did, because it did not preferentially export men. The coefficient, however, loses significance once we control for the victim's slave trade exposure.

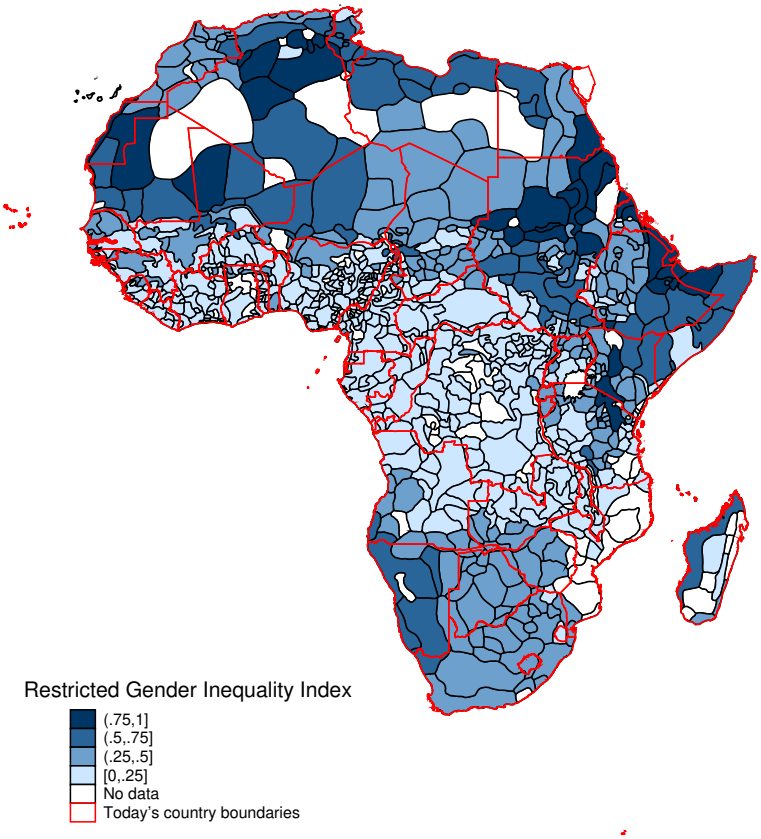
Table B.1: *The Slave Trade and Sexual Violence in Armed Conflict*

	Dependent variable: sexual violence (0-3)					
	(1)	(2)	(3)	(4)	(5)	(6)
The Slave Trade						
Transatlantic Slave Trade	-1.12**	-0.86*	-1.56***			
	(0.434)	(0.442)	(0.372)			
Indian Ocean Slave Trade				0.74***	0.74**	1.66
				(0.258)	(0.292)	(1.247)
Mean dep. var.	0.62	0.62	0.62	0.62	0.62	0.62
Conflict fixed effect	yes	yes	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes	yes	yes
Conflict-specific time trend		yes	yes		yes	yes
Victim's characteristic			yes			yes
Observations	900	900	893	900	900	893
Clusters	128	128	127	128	128	127
Adjusted R^2	0.277	0.341	0.343	0.273	0.340	0.277

Notes: OLS coefficient estimates of Equation 2.1. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. Explanatory variables are perpetrator-specific exposure to the Atlantic and the Indian slave trade, respectively. Both variables are constructed as: ln(1+Number of slaves/Ethnic group's land area) as in (Nunn and Wantchekon [2011]). All explanatory variables are normalized and range between 0 and 1. Standard errors are clustered at the perpetrator's level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

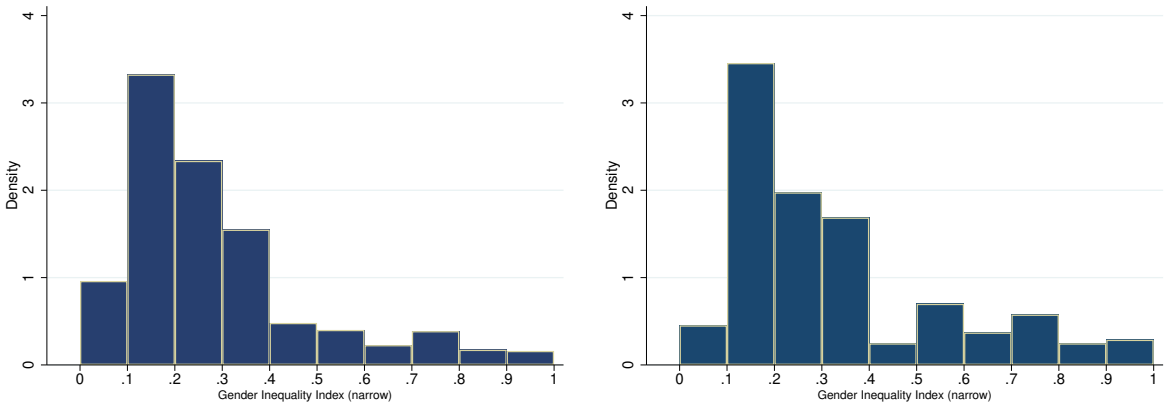
B.3 Additional Figures and Tables

Figure B.4: *Distribution of the Restricted eGII across Africa*



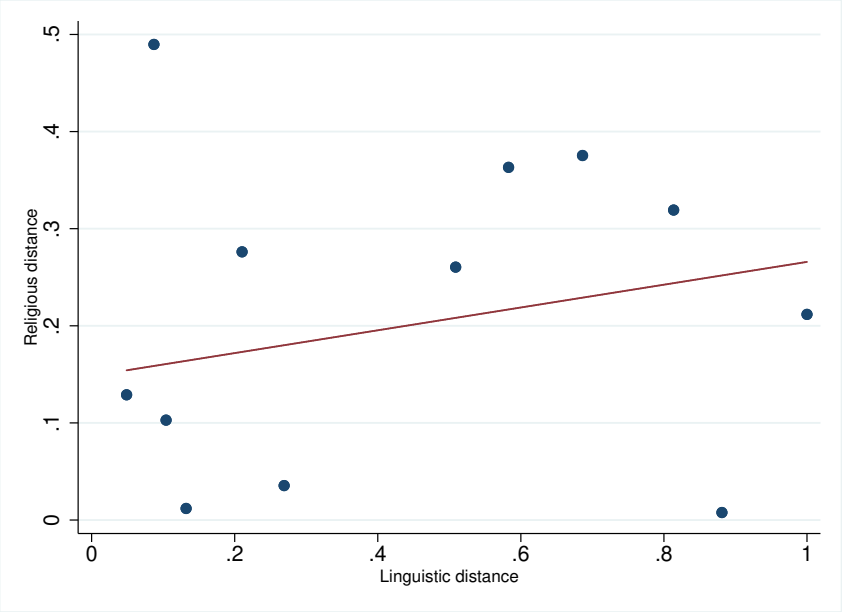
Notes: Restricted Gender Inequality Index across Murdock's ethnicities in Africa and contemporary country borders.

Figure B.5: *Distribution of the Restricted eGII*



Notes: Left: Distribution of the restricted eGII in Africa. Mean (standard deviation): 0.28 (0.20); right: Distribution of the restricted eGII in our sample. Mean (standard deviation): 0.32 (0.23)

Figure B.6: *Correlation between Linguistic Distance and Religious Distance*



Notes: Correlation between linguistic distance between the combatants and their religious distance for the sample of ethnicities involved in inter-ethnic conflict. Correlation coefficient: 0.23***. Sources: Ethnologue and EPR-ED dataset.

Table B.2: *Dataset Extract Illustrating the Dyadic Structure*

Country	**Year**	**Conflict ID**	**Perpetrator**	**Victim**	**Perpetrator's ethnicity**	**Victim's ethnicity**	**SVAC**
Chad	1994	288	Government of Chad	CSNPD	Zaghawa, Bideyat	Sara	2
Chad	1994	288	CSNPD	Government of Chad	Sara	Zaghawa, Bideyat	0

Table B.3: *Perpetrator's Ethnic Characteristics and Sexual Violence in Armed Conflict (I)*

	Dependent variable: sexual violence (0-3)					
	(1)	(2)	(3)	(4)	(5)	(6)
Lineage, Residence and Family Arrangements						
Matrilineal (weighted)	-0.78***	-0.85***	-0.66**			
	(0.219)	(0.283)	(0.326)			
Matrilineal				-0.77***	-0.84***	-0.66**
				(0.218)	(0.276)	(0.315)
Adjusted R^2	0.292	0.363	0.367	0.292	0.364	0.368
Patrilocal (weighted)	0.65***	0.70***	0.60**			
	(0.201)	(0.249)	(0.284)			
Patrilocal				0.64***	0.69***	0.59**
				(0.205)	(0.248)	(0.286)
Adjusted R^2	0.291	0.362	0.363	0.291	0.361	0.362
Stem (weighted)	-0.42	-0.47*	-0.40			
	(0.340)	(0.283)	(0.521)			
Stem				-0.50	-0.52*	-0.28
				(0.334)	(0.301)	(0.533)
Adjusted R^2	0.278	0.347	0.347	0.282	0.350	0.351
Mean dep. var.	0.62	0.62	0.62	0.62	0.62	0.62
Conflict fixed effect	yes	yes	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes	yes	yes
Conflict-specific time trend		yes	yes		yes	yes
Victim's characteristic			yes			yes
Observations	900	900	893	900	900	893
Clusters	128	128	127	128	128	127

Notes: OLS coefficient estimates of Equation 2.1. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. Explanatory variables are perpetrator-specific ethnic characteristics—either unweighted or weighted by the ethnic group's land area—capturing lineage systems (matrilineal), residence patterns (virilocal) and family arrangements (stem). All explanatory variables are normalized and range between 0 and 1. Standard errors are clustered at the perpetrator's level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table B.4: *Perpetrator's Ethnic Characteristics and Sexual Violence in Armed Conflict (II)*

	Dependent variable: sexual violence (0-3)					
	(1)	(2)	(3)	(4)	(5)	(6)
Subsistence Activities (I)						
Gathering (weighted)	1.70	1.40	1.55			
	(1.431)	(1.522)	(1.560)			
Gathering				0.71	0.27	0.82
				(1.823)	(2.000)	(2.246)
Adjusted R^2	0.274	0.340	0.341	0.271	0.338	0.340
Hunting (weighted)	1.31	1.38	2.77			
	(2.212)	(2.215)	(2.703)			
Hunting				2.50	2.46	3.40
				(2.275)	(2.294)	(2.764)
Adjusted R^2	0.272	0.340	0.343	0.276	0.343	0.345
Agriculture (weighted)	-1.38**	-1.34**	-1.47*			
	(0.569)	(0.605)	(0.877)			
Agriculture				-1.47***	-1.46**	-1.43*
				(0.545)	(0.578)	(0.786)
Adjusted R^2	0.295	0.362	0.363	0.300	0.367	0.368
Mean dep. var.	0.62	0.62	0.62	0.62	0.62	0.62
Conflict fixed effect	yes	yes	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes	yes	yes
Conflict-specific time trend		yes	yes		yes	yes
Victim's characteristic			yes			yes
Observations	900	900	893	900	900	893
Clusters	128	128	127	128	128	127

Notes: OLS coefficient estimates of Equation 2.1. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. Explanatory variables are perpetrator-specific ethnic characteristics—either unweighted or weighted by the ethnic group's land area—capturing dependence on different subsistence activities (gathering, hunting, agriculture). All explanatory variables are normalized and range between 0 and 1. Standard errors are clustered at the perpetrator's level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table B.5: *Perpetrator's Ethnic Characteristics and Sexual Violence in Armed Conflict (III)*

	Dependent variable: sexual violence (0-3)					
	(1)	(2)	(3)	(4)	(5)	(6)
Subsistence Activities (II)						
Plough (weighted)	0.15	0.07	-0.05			
	(0.606)	(0.701)	(0.701)			
Plough				0.31	-0.23	0.09
				(0.689)	(0.506)	(0.781)
Adjusted R^2	0.274	0.340	0.341	0.271	0.338	0.340
Husbandry (weighted)	1.50***	1.56***	1.56*			
	(0.567)	(0.578)	(0.812)			
Husbandry				1.65***	1.70***	1.59**
				(0.506)	(0.523)	(0.713)
Adjusted R^2	0.296	0.366	0.367	0.304	0.375	0.375
Pastoralism (weighted)	1.50***	1.56***	1.57*			
	(0.562)	(0.575)	(0.810)			
Pastoralism				1.67***	1.73***	1.62**
				(0.504)	(0.522)	(0.713)
Adjusted R^2	0.297	0.367	0.367	0.305	0.376	0.376
Mean dep. var.	0.62	0.62	0.62	0.62	0.62	0.62
Conflict fixed effect	yes	yes	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes	yes	yes
Conflict-specific time trend		yes	yes		yes	yes
Victim's characteristic			yes			yes
Observations	900	900	893	900	900	893
Clusters	128	128	127	128	128	127

Notes: OLS coefficient estimates of Equation 2.1. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. Explanatory variables are perpetrator-specific ethnic characteristics—either unweighted or weighted by the ethnic group's land area—capturing the use of the plough, and dependence on different subsistence activities (animal husbandry and pastoralism). The specification with "use of the plough" as an explanatory variable controls for dependence on agriculture. All explanatory variables are normalized and range between 0 and 1. Standard errors are clustered at the perpetrator's level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table B.6: *Perpetrator's eGII and Sexual Violence in Armed Conflict: Robustness*

	Dependent Variable: Sexual Violence (0-3)					
	Victim FE		Country FE		No time variation	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
	(1)	(2)	(3)	(4)	(5)	(6)
Ethnic Gender Inequality Index						
eGII	0.84	0.81*	1.47***	1.48***	1.54**	1.59***
	(0.550)	(0.472)	(0.461)	(0.450)	(0.630)	(0.585)
Adjusted R^2	0.547	0.548	0.296	0.298	0.260	0.269
Restricted eGII	1.20	1.07*	1.01**	1.10***	1.18**	1.40***
	(0.726)	(0.594)	(0.458)	(0.391)	(0.566)	(0.528)
Adjusted R^2	0.550	0.550	0.283	0.287	0.251	0.267
Conflict fixed effect	yes	yes			yes	yes
Country fixed effect			yes	yes		
Year fixed effect	yes	yes	yes	yes		
Conflict-specific time trend	yes	yes				
Country-specific time trend			yes	yes		
Victim fixed effect	yes	yes				
Observations	880	880	900	900	266	266
Clusters	127	127	128	128	128	128

Notes: OLS coefficient estimates of Equation 2.1. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. Explanatory variables are perpetrator-specific ethnic characteristics related to descent, residence patterns, family arrangements and subsistence activities. All explanatory variables are normalized and range between 0 and 1. Observations are at the perpetrator-victim-conflict level. Columns (1), (3) and (5) report coefficients for covariates weighted by the size of the ethnic group, while columns (2), (4) and (6) report coefficients for unweighted covariates. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table B.7: *Perpetrator's Ethnic Characteristics: Robustness (I)*

	Dependent Variable: Sexual Violence (0-3)					
	Victim FE		Country FE		No time variation	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
	(1)	(2)	(3)	(4)	(5)	(6)
Lineage, Residence and Family Arrangements						
Matrilineal	-0.33*	-0.35*	-0.84***	-0.82***	-0.58***	-0.55***
	(0.179)	(0.182)	(0.200)	(0.199)	(0.206)	(0.208)
Adjusted R^2	0.547	0.547	0.290	0.290	0.249	0.247
Virilocal	0.33*	0.31	0.83***	0.80***	0.51**	0.47**
	(0.182)	(0.192)	(0.186)	(0.185)	(0.213)	(0.216)
Adjusted R^2	0.547	0.546	0.299	0.296	0.244	0.242
Stem	0.06	0.42	-0.54**	-0.59**	-0.48	-0.71*
	(0.755)	(0.735)	(0.258)	(0.275)	(0.403)	(0.407)
Adjusted R^2	0.544	0.547	0.272	0.277	0.234	0.245
Conflict fixed effect	yes	yes			yes	yes
Country fixed effect			yes	yes		
Year fixed effect	yes	yes	yes	yes		
Conflict-specific time trend	yes	yes				
Country-specific time trend			yes	yes		
Victim fixed effect	yes	yes				
Mean dep. var.	0.62	0.62	0.62	0.62	0.57	0.57
Observations	880	880	900	900	266	266
Clusters	127	127	128	128	128	128

Notes: OLS coefficient estimates of Equation 2.1. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. Explanatory variables are perpetrator-specific ethnic characteristics related to descent, residence patterns, family arrangements, and subsistence activities. All explanatory variables are normalized and range between 0 and 1. Observations are at the perpetrator-victim-conflict level. Columns (1), (3) and (5) report coefficients for covariates weighted by the size of the ethnic group, while columns (2), (4) and (6) report coefficients for unweighted covariates. Standard errors are clustered at the perpetrator's level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table B.8: *Perpetrator's Ethnic Characteristics: Robustness (II)*

	Dependent Variable: Sexual Violence (0-3)					
	Victim FE		Country FE		No time variation	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
	(1)	(2)	(3)	(4)	(5)	(6)
Subsistence Activities (I)						
Gathering	2.91	3.37	0.09	-0.71	1.52	0.35
	(1.790)	(2.301)	(1.536)	(1.841)	(1.318)	(1.489)
Adjusted R^2	0.548	0.547	0.259	0.259	0.230	0.226
Hunting	4.80*	4.66*	1.55	2.18	0.91	2.00
	(2.432)	(2.460)	(2.110)	(2.242)	(2.376)	(2.446)
Adjusted R^2	0.550	0.549	0.261	0.264	0.226	0.230
Agriculture	-1.48*	-1.30*	-1.02**	- 1.07**	-1.26**	-1.50***
	(0.845)	(0.715)	(0.483)	(0.472)	(0.588)	(0.572)
Adjusted R^2	0.551	0.550	0.279	0.280	0.251	0.264
Mean dep. var.	0.62	0.62	0.62	0.62	0.57	0.57
Conflict fixed effect	yes	yes			yes	yes
Country fixed effect			yes	yes		
Year fixed effect	yes	yes	yes	yes		
Conflict-specific time trend	yes	yes				
Country-specific time trend			yes	yes		
Victim fixed effect	yes	yes				
Observations	880	880	900	900	266	266
Clusters	127	127	128	128	128	128

Notes: OLS coefficient estimates of Equation 2.1. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. Explanatory variables are perpetrator-specific ethnic characteristics related to descent, residence patterns, family arrangements, and subsistence activities. All explanatory variables are normalized and range between 0 and 1. Observations are at the perpetrator-victim-conflict level. Columns (1), (3) and (5) report coefficients for covariates weighted by the size of the ethnic group, while columns (2), (4) and (6) report coefficients for unweighted covariates. Standard errors are clustered at the perpetrator's level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table B.9: *Perpetrator's Ethnic Characteristics: Robustness (III)*

	Dependent Variable: Sexual Violence (0-3)					
	Victim FE		Country FE		No time variation	
	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted
	(1)	(2)	(3)	(4)	(5)	(6)
Subsistence Activities (II)						
Plough	-0.78***	-0.74***	0.21	0.29	-0.47	-0.47
	(0.248)	(0.272)	(0.480)	(0.481)	(0.296)	(0.304)
Adjusted R^2	0.555	0.554	0.278	0.281	0.229	0.229
Husbandry	1.21	1.14*	1.09**	1.18***	1.24*	1.48**
	(0.778)	(0.654)	(0.458)	(0.440)	(0.657)	(0.603)
Adjusted R^2	0.549	0.549	0.282	0.286	0.248	0.263
Pastoralism	1.248	1.171*	1.080**	1.186***	1.24*	1.50**
	(0.779)	(0.654)	(0.456)	(0.439)	(0.648)	(0.602)
Adjusted R^2	0.549	0.550	0.282	0.286	0.249	0.264
Mean dep. var.	0.62	0.62	0.62	0.62	0.57	0.57
Conflict fixed effect	yes	yes			yes	yes
Country fixed effect			yes	yes		
Year fixed effect	yes	yes	yes	yes		
Conflict-specific time trend	yes	yes				
Country-specific time trend			yes	yes		
Victim fixed effect	yes	yes				
Observations	880	880	900	900	266	266
Clusters	127	127	128	128	128	128

Notes: OLS coefficient estimates of Equation 2.1. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. Explanatory variables are perpetrator-specific ethnic characteristics related to descent, residence patterns, family arrangements, and subsistence activities. All explanatory variables are normalized and range between 0 and 1. Observations are at the perpetrator-victim-conflict level. Columns (1), (3) and (5) report coefficients for covariates weighted by the size of the ethnic group, while columns (2), (4) and (6) report coefficients for unweighted covariates. Standard errors are clustered at the perpetrator's level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table B.10: *Restricted eGII: PCA Loadings*

Variables	*Loading*
Gender Equal Traits	
Matrilineal	-0.29
Dependence on agriculture	-0.41
Gender Unequal Traits	
Virilocal	0.30
Dependence on pastoralism	0.57
Dependence on animal husbandry	0.57
Kaiser-Meyer-Olkin's measure of sampling adequacy	0.58

Notes: Loadings from the principal component analysis on the restricted eGII.

Table B.11: *Restricted Gender Inequality Index and Gender Inequality Outcomes (DHS)*

	Dependent variable				
				Intimate partner violence:	
	Female employment	Son Preference	Justifies beating	Physical	Sexual
	(1)	(2)	(3)	(4)	(5)
eGII (weighted)	-0.257***	0.054***	0.147***	-0.014	0.061*
	(0.051)	(0.013)	(0.048)	(0.026)	(0.034)
Adjusted R^2	0.092	0.046	0.156	0.018	0.026
eGII (unweighted)	-0.279***	0.059***	0.156***	-0.020	0.056
	(0.053)	(0.013)	(0.050)	(0.028)	(0.038)
Adjusted R^2	0.093	0.047	0.156	0.018	0.026
Mean dep. var.	0.580	0.032	0.534	0.064	0.100
Observations	571,184	428,718	481,728	113,192	69,706
Clusters	618	587	564	458	348
Countries	24	24	22	19	15
Years	27	25	20	15	11

Notes: Dependent variables: column (1): female employment; column (2): son preference, defined as (ideal number of boys - ideal number of girls)/(total number of wanted children); column (3) wife beating is justified in at least one of the following instances: she goes out without telling him, she neglects the children, she argues with him, she refuses to have sex with him, she burns the food; column (5) Faced at least one of the following severe physical violence events in the past 12 months: been kicked or dragged; been strangled; been threatened with knife/gun or other weapon; column (6) Faced at least one of the following sexual violence events in the past 12 months: physically forced into unwanted sex; forced into other unwanted sexual acts; physically forced to perform sexual acts she didn't want to. Explanatory variables: perpetrator's eGII weighted by the ethnic group land area and unweighted. All explanatory variables are normalized and range between 0 and 1. All columns include country and year fixed effects. Standard errors are clustered at the ethnic group's level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table B.12: *Restricted Inequality Index and Gender Attitudes (Afrobarometer)*

	Dependent variable					
	Men better political leaders (1)	Women and men equal rights (2)	Educating boys priority (3)	Men more right to a job (4)	Women right to own land (5)	Women care home and kids (6)
eGII (weighted)	0.152***	-0.128**	0.186***	0.090	-0.059	0.052
	(0.036)	(0.053)	(0.064)	(0.058)	(0.041)	(0.046)
Adjusted R^2	0.051	0.062	0.041	0.040	0.117	0.051
eGII (unweighted)	0.182***	-0.143**	0.198***	0.116*	-0.081*	0.046
	(0.041)	(0.060)	(0.072)	(0.062)	(0.045)	(0.049)
Adjusted R^2	0.051	0.062	0.040	0.040	0.117	0.052
Mean dep. var.	0.301	0.717	0.184	0.440	0.740	0.577
Observations	141,567	81,026	36,971	33,420	33,699	32,676
Clusters	770	638	473	413	412	413
Countries	34	34	32	31	31	31
Rounds	5	4	1	1	1	1

Notes: Dependent variables: column (1): agreeing with the statement "Men make better political leaders than women, and should be elected rather than women" as opposed to "Women should have the same chance of being elected to political office as men"; column (2) agreeing with the statement "In our country, women should have equal rights and receive the same treatment as men do" as opposed to "In our country, women should have equal rights and receive the same treatment as men do"; column (3): agreeing with the statement "If funds for schooling are limited, a boy should always receive an education in school before a girl" as opposed to "If funds for schooling are limited, a family should send the child with the greatest ability to learn"; column (4) agreeing with the statement "When jobs are scarce, men should have more right to a job than women"; column (5): agreeing with the statement "Women should have the same rights as men to own and inherit land"; column (6): agreeing with the statement "In general, it is better for a family if a woman has the main responsibility for taking care of the home and children rather than a man". Explanatory variables: perpetrator's eGII weighted by the ethnic group land area and unweighted. All explanatory variables are normalized and range between 0 and 1. All columns include country fixed effects. Columns (1) and (2) include survey round fixed effects. Standard errors are clustered at the ethnic group's level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table B.13: *Robustness: Controlling for Victim's Characteristics*

	Dep. Variable: Sexual Violence (0-3)			
	(1)	(2)	(3)	(4)
Victim's eGII	0.83	-1.14	1.13	
	(0.879)	(0.806)	(0.997)	
Perpetrator *more* unequal	2.77**		2.64**	2.17
	(1.057)		(1.047)	(2.101)
Perpetrator *less* unequal		-0.75	-0.49	0.88
		(0.658)	(0.641)	(1.527)
Conflict fixed effect	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes
Conflict-Specific time trends	yes	yes	yes	yes
Victim fixed effect				yes
Mean Dep. Var	0.62	0.62	0.62	0.62
Observations	643	643	643	625
Adjusted R^2	0.60	0.59	0.60	0.70

Notes: OLS coefficient estimates of Equations 2.3 and 2.4. The sample is restricted to include inter-ethnic conflicts only. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. The explanatory variables are the following: the victim's eGII (weighted by the ethnic group land area); the absolute distance in the eGII between perpetrator and victim when the perpetrator is more gender unequal than the victim; the absolute distance in the eGII between perpetrator and victim when the perpetrator is less gender unequal than the victim. Standard errors are clustered at the dyad level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table B.14: *Robustness: Abstracting from Temporal Variation*

	Dependent Variable: Sexual Violence (0-3)			
	(1)	(2)	(3)	(4)
Absolute distance ($\lvert eGII_p - eGII_v \rvert$)	1.98*			
	(1.027)			
Perpetrator's eGII		-0.19	2.71	1.36
		(0.888)	(1.753)	(1.916)
Perpetrator *more* unequal		1.95*		2.26**
		(1.007)		(0.966)
Perpetrator *less* unequal			1.60	2.11
			(1.745)	(1.854)
Conflict fixed effect	yes	yes	yes	yes
Mean Dep. Var	0.54	0.54	0.54	0.54
Observations	189	189	189	189
Adjusted R^2	0.226	0.245	0.234	0.256

Notes: OLS coefficient estimates of Equations 2.3 and 2.4. The sample is restricted to include inter-ethnic conflicts only. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. The explanatory variables are the following: the absolute distance in the eGII between perpetrator and victim; the perpetrator's eGII (weighted by the ethnic group land area); the absolute distance in the eGII between perpetrator and victim when the perpetrator is more gender unequal than the victim; the absolute distance in the eGII between perpetrator and victim when the perpetrator is less gender unequal than the victim. Robust standard errors are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table B.15: *Robustness: Alternative Fixed Effects and Alternative Versions of the eGII*

	Dependent Variable: Sexual Violence (0-3)			
	Conflict-year FE	Country FE	Unweighted eGII	Restricted eGII
	(1)	(2)	(3)	(4)
Perpetrator *more* unequal	1.18**	1.62***	1.14*	1.27**
	(0.521)	(0.514)	(0.618)	(0.502)
Perpetrator *less* unequal	1.60	1.45	0.81	1.55
	(1.067)	(1.246)	(1.239)	(1.260)
Conflict fixed effect	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes
Conflict-specific time trends		yes	yes	yes
Country fixed effect		yes		
Conflict-Year fixed effect	yes			
Perpetrator fixed effect	yes	yes	yes	yes
Mean Dep. Var	0.62	0.62	0.62	0.62
Observations	604	623	623	623
Adjusted R^2	0.64	0.74	0.74	0.74

Notes: OLS coefficient estimates of Equations 2.3 and 2.4. The sample is restricted to include inter-ethnic conflicts only. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. The explanatory variables are the following: the absolute distance in the eGII between perpetrator and victim when the perpetrator is more gender unequal than the victim; the absolute distance in the eGII between perpetrator and victim when the perpetrator is less gender unequal than the victim. Standard errors are clustered at the dyad level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table B.16: *Robustness: Assigning to Governments a Country-Level Measure of the eGII*

	Dependent Variable: Sexual Violence (0-3)				
	(1)	(2)	(3)	(4)	(5)
Absolute distance ($\lvert eGII_p - eGII_v \rvert$)	1.69***				
	(0.495)				
Perpetrator's eGII		0.59	1.70*	0.60	
		(0.832)	(0.886)	(0.934)	
Perpetrator *more* unequal		1.83**		1.83**	1.75***
		(0.867)		(0.867)	(0.434)
Perpetrator *less* unequal			-0.43	0.01	1.52
			(0.933)	(0.855)	(1.112)
Conflict fixed effect	yes	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes	yes
Conflict-specific time trends	yes	yes	yes	yes	yes
Perpetrator fixed effect	yes				yes
Mean dep. var.	0.62	0.62	0.62	0.62	0.62
Observations	633	653	653	653	633
Adjusted R^2	0.600	0.367	0.360	0.366	0.599

Notes: OLS coefficient estimates of Equations 2.3 and 2.4. The sample is restricted to include inter-ethnic conflicts only. Government forces' eGII is a country-level measure capturing the weighted average of ethnic groups' eGII within a country, weighted by the size of their land area. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. The explanatory variables are the following: the absolute distance in the eGII between perpetrator and victim; the absolute distance in the eGII between perpetrator and victim when the perpetrator is more gender unequal than the victim; the absolute distance in the eGII between perpetrator and victim when the perpetrator is less gender unequal than the victim. Standard errors are clustered at the dyad level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table B.17: *Robustness: Two-way Clustering*

	Dependent Variable: Sexual Violence (0-3)				
	(1)	(2)	(3)	(4)	(5)
Absolute distance ($\lvert eGII_p - eGII_v \rvert$)	1.53**				
	(0.623)				
Perpetrator's eGII		0.58	2.05	1.13	
		(0.468)	(1.398)	(1.379)	
Perpetrator *more* unequal		1.44**		1.51***	1.53**
		(0.584)		(0.569)	(0.650)
Perpetrator *less* unequal			0.20	0.64	1.56
			(1.328)	(1.350)	(1.488)
Conflict fixed effect	yes	yes	yes	yes	yes
Year fixed effect	yes	yes	yes	yes	yes
Conflict-specific time trends	yes	yes	yes	yes	yes
Perpetrator fixed effect	yes				yes
Mean dep. var.	0.62	0.62	0.62	0.62	0.62
Observations	623	643	643	643	623
Adjusted R^2	0.579	0.379	0.374	0.379	0.578

Notes: OLS coefficient estimates of Equations 2.3 and 2.4. The sample is restricted to include inter-ethnic conflicts only. The dependent variable is an index ranging between 0 and 3 that captures the intensity of sexual violence. The explanatory variables are the following: the absolute distance in the eGII between perpetrator and victim when the perpetrator is more gender unequal than the victim; the absolute distance in the eGII between perpetrator and victim when the perpetrator is less gender unequal than the victim. Standard errors are clustered at the perpetrator and victim level. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Appendix C

Appendix to Chapter 3

C.1 Data Sources

Ethnic groups: Data on ethnic groups comes from the Ethnic Power Relations (EPR) dataset family (Vogt *et al.* [2015]). EPR lists each politically relevant ethnic group for each country in each year, and the respective access (or lack thereof) to executive government power. EPR defines an ethnic group as an identity group that defines itself or is defined by others along linguistic, religious or racial characteristics.
Source: `https://icr.ethz.ch/data/epr/core/`

Linguistic distance: I retrieve data on languages (up to 3) spoken by each ethnic group from the Ethnic Dimensions Dataset. I compute a measure of linguistic distance (called cladistic distance) using information on linguistic trees coming from the Ethnologue database. I describe the methodology for the construction of the linguistic distance measure in section 3.3.
Sources: `https://icr.ethz.ch/data/epr/ed/` and `https://www.ethnologue.com/browse/names`

Conflict: Data on ethnic civil conflicts comes from the UCDP/PRIO Armed Conflict Dataset (version 20.1), with information on the identity of the combatants involved in a conflict (rebel groups and governments). Information on the ethnic identity of rebel groups comes from the ACD2EPR dataset.
Sources: `https://ucdp.uu.se/downloads/` and `https://icr.ethz.ch/data/epr/acd2epr/`

Geodesic distance: I compute geodesic distances between each group's centroid, which I calculate based on ethnic settlements provided by the GeoEPR dataset (Wucherpfennig *et al.* [2011]). Throughout the analysis, I use log geodesic distance.
Source: `https://icr.ethz.ch/data/epr/geoepr/`

Absolute difference in elevation: I combine GeoEPR with data on elevation at the grid level (2.5 arc-minute resolution) provided by the Worldclim Global Climate Database (Fick and Hijmans [2017]). I measure average elevation of each ethnic group, and then calculate the absolute distance in elevation between each ethnic group and the ethnicities forming the government.
Source: `https://www.worldclim.org/data/worldclim21.html`

Absolute difference in ruggedness: I produce a measure of ruggedness for each ethnic group using the standard deviation of the Worldclim elevation data. I calculate the absolute distance in ruggedness between each ethnic group and the ethnicities forming the government.
Source: `https://www.worldclim.org/data/worldclim21.html`

Absolute difference in the Caloric Suitability Index: Data on the potential agricultural output come from the Caloric Suitability Index (CSI) provided by Galor and Özak [2016]. This measure reflects the potential caloric yield of a grid cell based on the Global Agro-Ecological Zones (GAEZ) project of the Food and Agriculture Organization (FAO). The Index results from a combination of climatic and grographic variables unaffected by human activity. I use the pre-1500 average CSI measure that includes cells with zero productivity. I calculate the average CSI for each ethnic group, and calculate the absolute distance in CSI between each ethnic group and the ethnicities forming the government.
Source: `https://ozak.github.io/Caloric-Suitability-Index/`

Absolute difference in temperature: Temperature data comes from the WorldClim Global Climate Database. I compute a time-varying measure of mean temperature for each group, and calculate the absolute distance in mean temperature between each ethnic group and the ethnicities forming the government.
Source: `https://www.worldclim.org/data/worldclim21.html`

Absolute difference in rainfall: Rainfall data comes from the WorldClim Global Climate Database. I compute a time-varying measure of average rainfall for each group, and calculate the absolute distance in mean rainfall between each ethnic group and the ethnicities forming the government.
Source: `https://www.worldclim.org/data/worldclim21.html`

Ancestral homelands of contemporary ethnicities: I retrieve the geolocation of ethnic groups' ancestral homelands from the Murdock Map, digitized by Nunn and Wantchekon [2011]. I merge EPR ethnicities to the corresponding Murdock groups' counterparts using the LEDA R-package (see details in Section C.2.)
Source: `https://scholar.harvard.edu/nunn/pages/data-0`

The Bantu expansion route. Data on the Bantu expansion route comes from work by Grollemund *et al.* [2015]. They reconstructed the route by collecting the indigenous geographic location of 424 Bantu languages, including now-extinct languages. Using Bayesian techniques on a sample of 100 lexical items, they construct a pyhlogenetic tree connecting all these languages. Based on this, they reconstruct the probable ancestral geographical locations of each of the internal nodes of the phylogenetic tree through a model calibrated using archaeological evidence (e.g., node 1 is dated back to 4000-5000 before present in the Grassfields region of Cameroon, on the basis of an archaeological site called Shum Laka, the principal site associated with Bantu homelands). Using other archeological sites, they calibrate other branching points of the tree. using a Brownian motion model, they infer the ancestral latitude and longitude for each internal node of the tree, which they then connect using straight lines. Additional methodological details can be found in Grollemund *et al.* [2015] and in the article's Supplementary Appendix.

Opinions on government performance: Data on individuals' opinion on the performance of the current government comes from seven rounds of the Afrobarometer survey, conducted between 1999 and 2017 in 27 African countries. See Table C.26 for additional details on the questions asked in each round. Refer to Section 3.8 for details on how I construct a measure of an individual's degree of disagreement with policies implemented by the current government. I merge the self-reported ethnolinguistic affiliation of Afrobarometer respondents ethnicities to the corresponding EPR groups using the LEDA R-package (see details in Section C.2.)
Source: `http://afrobarometer.org/data`

Age: Age of respondent at the time of survey.

Gender: An indicator variable equal to one if a respondent is female.

Rural: An indicator variable for rural locations.

C.2 Merging Ethnic Groups Across Datasets

To merge ethnicities across datasets, I use a recent R-package created by Müller-Crepon *et al.* [Forthcoming] called Linking Ethnic Data for Africa (LEDA). Since ethnic categories vary considerably across datasets, Müller-Crepon *et al.* [Forthcoming] created an algorithm that systematically links ethnic groups across 11 dataset. Using a dictionary-based linking procedure, they match more than 8,100 ethnicities via the list of known language families, languages, and dialects from the 16th edition of the Ethnologue database.

Since links are formed through the linguistic tree, the LEDA package allows the researcher to choose the level of precision for each match. One option is to use the so-called *set overlap* rule, which generates a link between any two groups that share at least one language node at a specified level of the language tree. The higher the level specified, the higher is the completeness, but the lower the precision of the match. Throughout the analysis, I opt for precision of matches, and choose the *dialect* level (i.e., the lowest level) of the linguistic tree of ethnic group in dataset A for generating matches to the ethnic group(s) in dataset B.

Even after choosing the matching procedure that favors precision, some matches remain more precise than others. Conveniently, for each match, the package provides an indication of the degree of precision by indicating the extent to which the linguistic tree of an ethnic group in dataset A overlaps with the linguistic tree of an ethnic group in dataset B. Whenever LEDA matches an ethnic group in dataset A (e.g., the EPR dataset) to multiple groups in dataset B (e.g., the Murdock Map), I employ this accuracy measure as a weight when calculating statistics for group in dataset A based on multiple groups in dataset B (e.g., when computing a measure of Bantu exposure for each EPR group based on the Bantu Index of multiple Murdock groups). Groups in dataset B that constitute more accurate matches will get more weight than groups constituting weaker links.

For additional information on the LEDA package, see: `http://www.carlmueller-crepon.org/project/ethnic_matching/`

C.3 Summary Statistics

Table C.1: *Number of Government Changes by Country*

Country	Number of changes	Country	Number of changes
Angola	0	Malawi	1
Benin	8	Mali	1
Botswana	0	Mauritania	0
Burundi	0	Mauritius	8
Cameroon	2	Morocco	0
Central African Republic	5	Mozambique	1
Chad	8	Namibia	1
Comoros	0	Niger	10
Congo	9	Nigeria	9
Congo, DRC	8	Senegal	2
Cote d'Ivoire	5	Sierra Leone	8
Djibouti	2	South Africa	1
Egypt	0	South Sudan	0
Equatorial Guinea	0	Sudan	2
Eritrea	1	Tanzania	1
Ethiopia	2	The Gambia	0
Gabon	2	Togo	5
Ghana	6	Uganda	5
Guinea	5	Zambia	1
Guinea-Bissau	6	Zimbabwe	4
Kenya	6		
Liberia	3		
Libya	0	**Total**	**138**

Notes: The table reports the number of changes in the ethnic identity of governments experienced by each country in the sample over the period 1961-2017.

Table C.2: *Summary Statistics: Africa*

	Mean (1)	Std. Dev. (2)	Min (3)	Max (4)	Obs. (5)
Conflict Variables					
Ethnic conflict	0.073	0.259	0	1	9,827
Conflict over government	0.049	0.215	0	1	9,827
Conflict over territory	0.025	0.155	0	1	9,827
Linguistic Distance					
Linguistic Distance^{W}	0.402	0.321	0	1	9,827
Linguistic Distance^{UW}	0.405	0.317	0	1	9,827
Geographic Controls					
Log geodesic distance^{W}	5.530	1.272	0	7.284	9,827
Log geodesic distance^{UW}	5.524	1.287	0	7.274	9,827
Absolute distance ruggedness^{W}	131.0	153.8	0	1,223	9,827
Absolute distance ruggedness^{UW}	133.5	155.4	0	1,238	9,827
Absolute distance elevation^{W}	214.9	242.4	0	1,442	9,827
Absolute distance elevation^{UW}	215.0	240.5	0	1,384	9,827
Absolute distance CSI^{W}	588.5	1,157	0	9,209	9,827
Absolute distance CSI^{UW}	589.9	1,135	0	9,206	9,827
Climatic Controls					
Average precipitation	92.27	61.62	0.280	353.3	9,827
Absolute distance precipitation^{W}	22.83	24.15	0	254.9	9,827
Absolute distance precipitation^{UW}	22.66	23.64	0	254.5	9,827
Average temperature	24.13	2.977	15.29	29.35	9,827
Absolute distance temperature^{W}	1.256	1.247	0	7.545	9,827
Absolute distance temperature^{UW}	1.257	1.235	0	7.516	9,827

Notes: The sample includes 236 distinct ethnic groups in 43 African countries over a period of 57 years (1961-2017). The superscript W indicates average distances between potential rebels and each ethnic group in the government coalition weighted by the role of ethnic groups in power (where senior partners receive double the weight of each junior partner); the superscript UW indicates unweighted average distances between potential rebels and each ethnic group in the government coalition.

Table C.3: *Summary Statistics: Split Ethnic Groups*

	Mean (1)	Std. Dev. (2)	Min (3)	Max (4)	Obs. (5)
Conflict Variables					
Ethnic conflict	0.078	0.268	0	1	7,874
Conflict over government	0.051	0.219	0	1	7,874
Conflict over territory	0.026	0.160	0	1	7,874
Linguistic Distance					
Linguistic Distance^{W}	0.448	0.327	0	1	7,874
Linguistic Distance^{UW}	0.457	0.325	0	1	7,874
Geographic Controls					
Log geodesic distance^{W}	5.753	1.232	0	7.254	7,874
Log geodesic distance^{UW}	5.784	1.222	0	7.254	7,874
Absolute distance ruggedness^{W}	111.9	129.7	0	1,192	7,874
Absolute distance ruggedness^{UW}	116.5	133.0	0	1,192	7,874
Absolute distance elevation^{W}	178.0	166.0	0	1,196	7,874
Absolute distance elevation^{UW}	180.2	164.0	0	1,196	7,874
Absolute distance CSI^{W}	420.2	688.1	0	6,654	7,874
Absolute distance CSI^{UW}	446.1	711.9	0	6,654	7,874
Climatic Controls					
Average precipitation	51.89	45.46	0.295	288.9	7,874
Absolute distance precipitation^{W}	16.76	15.71	0	101.7	7,874
Absolute distance precipitation^{UW}	17.17	15.87	0	97.09	7,874
Average temperature	24.39	3.294	17.08	29.35	7,874
Absolute distance temperature^{W}	1.159	1.040	0	6.576	7,874
Absolute distance temperature^{UW}	1.173	1.031	0	6.576	7,874

Notes: The sample includes 94 distinct ethnic groups in the Murdock Ethnographic Atlas split in 32 African countries observed over a period of 57 years (1961-2017). The superscript W indicates average distances between potential rebels and each ethnic group in the government coalition weighted by the role of ethnic groups in power (where senior partners receive double the weight of each junior partner); the superscript UW indicates unweighted average distances between potential rebels and each ethnic group in the government coalition.

Table C.4: *Summary Statistics: Bantu Region*

	Mean (1)	Std. Dev. (2)	Min (3)	Max (4)	Obs. (5)
Conflict Variables					
Ethnic conflict	0.060	0.238	0	1	3,628
Conflict over government	0.052	0.222	0	1	3,628
Conflict over territory	0.008	0.089	0	1	3,628
Linguistic Distance					
Linguistic Distance^{W}	0.257	0.273	0	1	3,628
Linguistic Distance^{UW}	0.264	0.274	0	1	3,628
Bantu Distance					
Bantu Distance^{W}	0.048	0.048	0	0.435	3,628
Bantu Distance^{UW}	0.054	0.274	0	0.596	3,628
Geographic Controls					
Log geodesic distance^{W}	5.603	1.349	0	7.284	3,628
Log geodesic distance^{UW}	5.600	1.349	0	7.274	3,628
Absolute distance ruggedness^{W}	148.1	120.7	0	676.1	3,628
Absolute distance ruggedness^{UW}	151.8	126.9	0	688.7	3,628
Absolute distance elevation^{W}	265.7	245.1	0	1,442	3,628
Absolute distance elevation^{UW}	269.8	243.2	0	1,384	3,628
Absolute distance CSI^{W}	331.7	597.6	0	3,646	3,628
Absolute distance CSI^{UW}	360.0	646.4	0	3,440	3,628
Climatic Controls					
Average precipitation	106.7	47.46	10.94	304.1	3,628
Absolute distance precipitation^{W}	19.41	16.57	0	101.7	3,628
Absolute distance precipitation^{UW}	19.42	16.29	0	97.09	3,628
Average temperature	22.47	2.319	15.29	25.96	3,628
Absolute distance temperature^{W}	1.312	1.079	0	5.911	3,628
Absolute distance temperature^{UW}	1.327	1.057	0	5.757	3,628

Notes: The sample includes 86 distinct ethnic groups in 17 African countries corresponding to regions covered by Bantu languages according to the Guthrie classification (Angola, Botswana, Burundi, Cameroon, Congo, Congo DRC, Equatorial Guinea, Gabon, Kenya, Malawi, Mozambique, Rwanda, South Africa, Tanzania, Uganda, Zambia, Zimbabwe) over a period of 57 years (1961-2017). The superscript W indicates average distances between potential rebels and each ethnic group in the government coalition weighted by the role of ethnic groups in power (where senior partners receive double the weight of each junior partner); the superscript UW indicates unweighted average distances between potential rebels and each ethnic group in the government coalition.

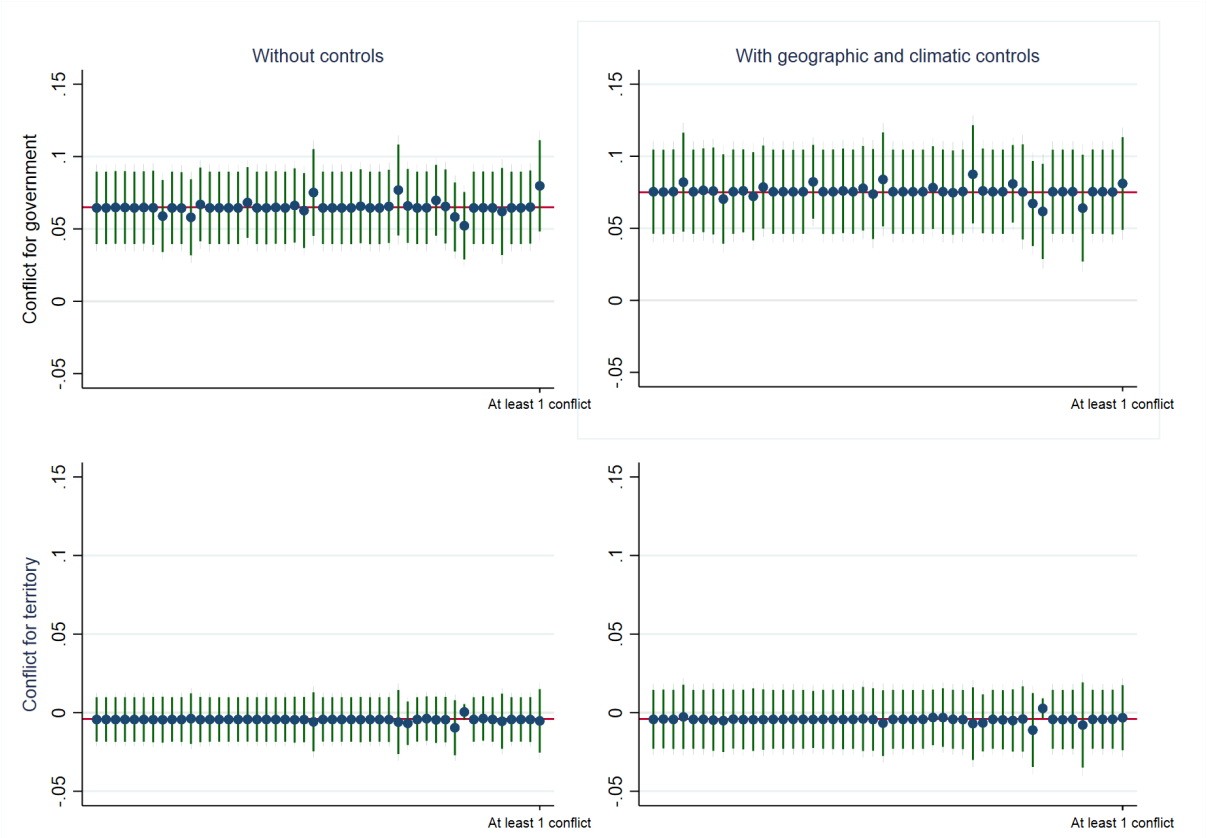
Table C.5: *Summary Statistics: Individual-Level Data*

	Mean (1)	Std. Dev. (2)	Min (3)	Max (4)	Obs. (5)
Opinion on government performance					
Bad opinion (first PC)	0.567	0.307	0	1	83,482
Bad opinion (average)	0.555	0.313	0	1	90,481
Linguistic Distance					
Linguistic distance^W	0.345	0.253	0	1	90,481
Individual Controls					
Age	36.65	14.48	15	105	90,481
Female	0.499	0.500	0	1	90,481
Rural	0.562	0.496	0	1	90,481
Geographic Controls					
Log geodesic distance^W	5.430	1.123	0	6.974	90,481
Absolute distance ruggedness^W	98.21	100.7	0	1223	90,481
Absolute distance CSI^W	314.8	461.4	0	6692	90,481
Absolute distance elevation^W	147.5	155.3	0	1214	90,481
Climatic Controls					
Average precipitation	83.66	43.14	3.618	330.7	90,481
Absolute distance precipitation^W	20.15	18.93	0	172.6	90,481
Average temperature	23.59	3.644	15.29	29.13	90,481
Absolute distance temperature^W	0.898	0.857	0	6.429	90,481

Notes: The sample includes individuals from 95 distinct ethnic groups in 27 African countries surveyed between 1999 and 2017. The superscript W indicates average distances between potential rebels and each ethnic group in the government coalition weighted by the role of ethnic groups in power (where senior partners receive double the weight of each junior partner).

C.4 Additional Figures and Tables

Figure C.1: *Checking for Outliers*



Notes: The figures show the stability of the estimates when dropping one country at a time from the sample, or when keeping only the 29 countries that experienced at least one conflict in the whole period (last coefficient on the right in each graph). The red horizontal line indicates the baseline coefficient; green and gray vertical lines indicate 90- and 95-percent confidence intervals, respectively. All specifications include country, year, country-year, and ethnicity fixed effects as well as ethnicity-specific time trends.

Table C.6: *Robustness: Controlling for Lagged Conflict Instrumenting the First Lag*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.027**	0.028*			-0.004	-0.004		
	(0.012)	(0.015)			(0.007)	(0.009)		
Linguistic distance^{UW}			0.025*	0.026			-0.008	-0.009
			(0.014)	(0.017)			(0.009)	(0.012)
Conflict$_{t-1}$	0.649***	0.649***	0.649***	0.649***	0.752***	0.752***	0.752***	0.751***
	(0.117)	(0.118)	(0.118)	(0.118)	(0.181)	(0.182)	(0.181)	(0.182)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,314	9,314	9,314	9,314	9,314	9,314	9,314	9,314

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1 -4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). Conflict$_{t-1}$ is a binary variable that is equal to 1 if an ethnic group was involved in a conflict over power (columns 1-4) or over territory (columns 5 -8) in the previous year. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 55 years (1963-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.7: *Robustness: Controlling for Potential Rebels Having Lost Power*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.061***	0.073***			-0.004	-0.004		
	(0.014)	(0.017)			(0.009)	(0.011)		
Linguistic distance^{UW}			0.057***	0.069***			-0.009	-0.011
			(0.014)	(0.016)			(0.012)	(0.014)
Ethnic group lost power	0.091*	0.090*	0.091*	0.090*	-0.006	-0.006	-0.004	-0.005
	(0.052)	(0.051)	(0.051)	(0.052)	(0.006)	(0.007)	(0.007)	(0.009)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.499	0.499	0.498	0.499	0.593	0.593	0.593	0.593

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1 -4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5 -8). *Ethnic group lost power* is equal to 1 if an ethnicity has lost government power. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.8: *Robustness: Controlling for Potential Rebels Having Gained Power*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.062***	0.073***			-0.004	-0.004		
	(0.015)	(0.018)			(0.008)	(0.011)		
Linguistic distance^{UW}			0.058***	0.070***			-0.010	-0.011
			(0.015)	(0.017)			(0.011)	(0.014)
Ethnic group gained power	-0.018	-0.017	-0.017	-0.016	-0.000	-0.001	-0.001	-0.002
	(0.014)	(0.014)	(0.014)	(0.014)	(0.005)	(0.005)	(0.005)	(0.005)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.498	0.498	0.498	0.499	0.593	0.593	0.593	0.593

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). *Ethnic group lost power* is equal to 1 if an ethnicity has gained government power. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.9: *Robustness: Alternative Linguistic Distance Measure*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.043**	0.057**			-0.008	-0.008		
	(0.018)	(0.024)			(0.007)	(0.010)		
Linguistic distance^{UW}			0.039**	0.049**			-0.015	-0.016
			(0.017)	(0.023)			(0.009)	(0.014)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.488	0.489	0.488	0.489	0.587	0.587	0.587	0.591

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). Linguistic distance indicates the (weighted or unweighted) linguistic distance between potential rebels and each ethnic group in the government coalition excluding the potential rebel itself from the computation if the latter is represented in the government. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.10: *Robustness: Only Countries that Experienced at Least One Conflict*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.080***	0.081***			-0.005	-0.003		
	(0.019)	(0.019)			(0.012)	(0.012)		
Linguistic distance^{UW}			0.074***	0.077***			-0.013	-0.012
			(0.018)	(0.017)			(0.016)	(0.016)
Mean of dep. var.	0.070	0.070	0.070	0.070	0.036	0.036	0.036	0.036
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	6,737	6,737	6,737	6,737	6,737	6,737	6,737	6,737
Adjusted R-squared	0.484	0.489	0.483	0.489	0.587	0.588	0.587	0.588

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). Linguistic distance indicates the (weighted or unweighted) linguistic distance between potential rebels and each ethnic group in the government coalition. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 28 African countries, 57 years (1961-2017), and 166 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.11: *Robustness: Only Countries that Experienced at Least One Government Change*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.068***	0.074***			-0.012	-0.012		
	(0.018)	(0.019)			(0.014)	(0.016)		
Linguistic distance^{UW}			0.063***	0.068***			-0.018	-0.019
			(0.018)	(0.018)			(0.017)	(0.018)
Mean of dep. var.	0.053	0.053	0.053	0.053	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	8,132	8,132	8,132	8,132	8,132	8,132	8,132	8,132
Adjusted R-squared	0.481	0.482	0.481	0.482	0.586	0.587	0.587	0.587

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). Linguistic distance indicates the (weighted or unweighted) linguistic distance between potential rebels and each ethnic group in the government coalition. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 31 African countries, 57 years (1961-2017), and 202 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.12: *Robustness: No Time Trends*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.023***	0.025***			-0.002	-0.003		
	(0.006)	(0.009)			(0.004)	(0.005)		
Linguistic distance^{UW}			0.020**	0.023*			-0.003	-0.005
			(0.008)	(0.011)			(0.006)	(0.007)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend								
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.429	0.429	0.429	0.430	0.487	0.488	0.487	0.488

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.13: *Robustness: Alternative Clustering (Country Level)*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.065***	0.075***			-0.004	-0.004		
	(0.016)	(0.017)			(0.009)	(0.011)		
Linguistic distance^{UW}			0.061***	0.072***			-0.009	-0.011
			(0.015)	(0.016)			(0.012)	(0.014)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.498	0.491	0.491	0.492	0.587	0.587	0.587	0.587

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Robust standard errors clustered at the country level are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.14: *Robustness: Alternative Clustering (Country-by-year Level)*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.065***	0.075***			-0.004	-0.004		
	(0.014)	(0.016)			(0.004)	(0.006)		
Linguistic distance^{UW}			0.061***	0.072***			-0.009*	-0.011
			(0.013)	(0.016)			(0.006)	(0.007)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.493	0.493	0.492	0.493	0.589	0.589	0.589	0.589

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Robust standard errors clustered at the country-year level are reported in parenthesis (corresponding to 2,000 clusters). *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.15: *Robustness: Alternative Clustering (Ethnic Group Level)*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.065***	0.075**			-0.004	-0.004		
	(0.024)	(0.034)			(0.008)	(0.012)		
Linguistic distance^{UW}			0.061**	0.072*			-0.009	-0.011
			(0.028)	(0.038)			(0.009)	(0.012)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.495	0.495	0.495	0.496	0.590	0.591	0.591	0.591

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Robust standard errors clustered at the ethnicity level are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.16: *Robustness: Lagged Linguistic Distance*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance$^{W}{}_{t-1}$	0.042***	0.039**			-0.002	-0.002		
	(0.015)	(0.018)			(0.009)	(0.012)		
Linguistic distance$^{UW}{}_{t-1}$			0.038**	0.037*			-0.006	-0.007
			(0.015)	(0.018)			(0.013)	(0.015)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,565	9,565	9,565	9,565	9,565	9,565	9,565	9,565
Adjusted R-squared	0.499	0.500	0.499	0.500	0.595	0.595	0.595	0.596

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). *Linguistic distance* indicates the (weighted or unweighted) lagged linguistic distance between potential rebels and the government. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 207 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.17: *Robustness: Balanced Panel*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.051**	0.052**			-0.006	-0.009		
	(0.019)	(0.022)			(0.011)	(0.018)		
Linguistic distance^{UW}			0.048**	0.047**			-0.013	-0.020
			(0.019)	(0.020)			(0.015)	(0.020)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	8,703	8,703	8,703	8,703	8,703	8,703	8,703	8,703
Adjusted R-squared	0.520	0.521	0.520	0.521	0.591	0.591	0.591	0.592

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 43 African countries, 57 years (1961-2017), and 207 distinct ethnic groups observed uninterruptedly throughout the period. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.18: *DID: Leads and Lags*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance$_{t-1}$		0.037	-0.030	-0.025		-0.026	0.002	0.008
		(0.026)	(0.038)	(0.041)		(0.019)	(0.010)	(0.010)
Linguistic distance$_{t}$	0.102**		0.108*	0.135**	-0.036*		-0.046**	-0.050**
	(0.039)		(0.059)	(0.064)	(0.019)		(0.019)	(0.021)
Linguistic distance$_{t+1}$				-0.036				0.003
				(0.022)				(0.020)
Mean of dep. var.	0.051	0.051	0.051	0.053	0.026	0.027	0.027	0.027
Country-ethn. fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls	yes	yes	yes	yes	yes	yes	yes	yes
Climatic controls	yes	yes	yes	yes	yes	yes	yes	yes
Observations	7,874	7,592	7,592	7,053	7,874	7,592	7,592	7,053
Adjusted R-squared	0.860	0.867	0.869	0.881	0.799	0.807	0.808	0.817

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government over power (columns 1-5) or over territory (columns 6-10) and 0 otherwise. *Linguistic distance* captures linguistic distance between each potential rebel and the ethnic groups at the government, either in period t-1, t, or t+1. *Linguistic distance* is standardized. For a description of geographic and climatic controls, refer to the notes in Table 3.2. The sample includes 97 unique ethnic groups split in 32 African countries in 56 years. Standard errors clustered at the ethnic group-country level in parenthesis (corresponding to 201 clusters). *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.19: *DID Robustness: Controlling for Lagged Conflict*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.058	0.077*			-0.025*	-0.041**		
	(0.037)	(0.041)			(0.015)	(0.020)		
Linguistic distance^{UW}			0.056	0.079*			-0.027*	-0.039**
			(0.040)	(0.044)			(0.015)	(0.018)
Conflict$_{t-1}$	0.396***	0.391***	0.395***	0.389***	0.150*	0.156*	0.151*	0.156*
	(0.051)	(0.050)	(0.052)	(0.051)	(0.083)	(0.084)	(0.083)	(0.084)
Mean of dep. var.	0.051	0.051	0.051	0.051	0.027	0.027	0.027	0.027
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	7,592	7,592	7,592	7,592	7,592	7,592	7,592	7,592
Adjusted R-squared	0.890	0.892	0.890	0.892	0.807	0.813	0.807	0.813

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). Conflict$_{t-1}$ is a binary variable that is equal to 1 if an ethnic group was involved in a conflict over power (columns 1-4) or over territory (columns 5-8) in the previous year. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 32 African countries, 57 years (1961-2017), and 94 distinct ethnic groups. Standard errors clustered at the ethnic group-country level in parenthesis (corresponding to 201 clusters). *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.20: *DID Robustness: Controlling for Being Part of Government Coalition*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.077**	0.101**			-0.027*	-0.037*		
	(0.036)	(0.039)			(0.014)	(0.019)		
Linguistic distance^{UW}			0.086**	0.120***			-0.029**	-0.035**
			(0.038)	(0.038)			(0.014)	(0.018)
Ethnic group is part of coalition	0.015	-0.011	0.018	-0.011	-0.012	-0.007	-0.013	-0.007
	(0.021)	(0.021)	(0.022)	(0.022)	(0.009)	(0.009)	(0.009)	(0.009)
Mean of dep. var.	0.051	0.051	0.051	0.051	0.026	0.026	0.026	0.026
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	7,874	7,874	7,874	7,874	7,874	7,874	7,874	7,874
Adjusted R-squared	0.858	0.860	0.858	0.861	0.794	0.799	0.794	0.799

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). Ethnic groups is part of coalition is equal to 1 if an ethnicity is represented in the government coalition and 0 otherwise. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 32 African countries, 57 years (1961-2017), and 94 distinct ethnic groups. Standard errors clustered at the ethnic group-country level in parenthesis (corresponding to 201 clusters). *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.21: *IV: Inference with Lee et al. [2020] tF Procedure*

	Conflict for government		Conflict for territory	
	(1)	(2)	(3)	(4)
Linguistic distance	0.191***	0.139***	-0.008	-0.010**
tF 0.05 se	(0.034)	(0.027)	(0.007)	(0.004)
Mean of dep. var	0.052	0.052	0.008	0.008
Distance type	w	uw	w	uw
Country-year FE	yes	yes	yes	yes
Ethnicity FE & trends	yes	yes	yes	yes
Controls	yes	yes	yes	yes
Observations	3,628	3,628	3,628	3,628

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government over power (columns 1-2) or over territory (columns 3-4) and zero otherwise. *Linguistic distance* captures linguistic distance between each potential rebel and the ethnic groups at the government. For a description of geographic controls, refer to the notes in Table 3.2. The sample includes 17 African countries in the Bantu region, 57 years (1961-2017), and 86 distinct ethnic groups. Two-way clustered standard errors by year and country, adjusted for the low number of clusters using the number of countries, are reported in parenthesis and are further re-adjusted using Lee *et al.* [2020] *tF* procedure. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.22: *IV: Bantu Exposure, Cultural Distance and Ethnic Civil Conflict*

Panel A	Dependent variable: Any ethnic conflict			
	OLS		2SLS	
	(1)	(2)	(3)	(4)
Linguistic distance	0.072***	0.068***	0.183***	0.129***
	(0.018)	(0.013)	(0.023)	(0.024)
Mean of dep. var	0.060	0.060	0.060	0.060
Panel B			Reduced Form	
Bantu distance			0.067***	0.060***
			(0.013)	(0.012)
Mean of dep. var			0.060	0.060
Adjusted R-squared			0.495	0.495
Panel C			First Stage	
			Linguistic distance	
Bantu distance			0.365***	0.468***
			(0.083)	(0.056)
Kleibergen-Paap F-statistic			14.62	53.64
Distance type	w	uw	w	uw
Country-year fixed effects	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes
Ethnicity year trends	yes	yes	yes	yes
Geographic controls	yes	yes	yes	yes
Climatic controls	yes	yes	yes	yes
Observations	3,628	3,628	3,628	3,628

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government and 0 otherwise. *Linguistic distance* captures linguistic distance between each potential rebel and the ethnic groups at the government. *Bantu distance* denotes the absolute difference in the exposure to the Bantu expansion between potential rebels and the government. Linguistic distance and Bantu distance are standardized. For a description of geographic controls, refer to the notes in Table 3.2. The sample includes 17 African countries in the Bantu region, 57 years (1961-2017), and 86 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis, and are adjusted for the low number of clusters using the number of countries. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.23: *IV Robustness: Controlling for Lagged Conflict*

	Ethnic conflict over power			Ethnic conflict over territory		
	OLS	2SLS		OLS	2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Linguistic distance	0.046***	0.129***	0.087***	0.000	-0.002	-0.002
	(0.011)	(0.010)	(0.013)	(0.008)	(0.008)	(0.008)
Conflict_{t-1}	0.446***	0.434***	0.439***	0.164	0.164	0.163
	(0.052)	(0.054)	(0.053)	(0.193)	(0.219)	(0.219)
Mean of dep. var.	0.052	0.052	0.052	0.008	0.008	0.008
Distance type	w	w	uw	w	w	uw
Country-year fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes
Geographic controls	yes	yes	yes	yes	yes	yes
Climatic controls	yes	yes	yes	yes	yes	yes
Observations	3,537	3,537	3,537	3,537	3,537	3,537

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government over power (columns 1-3) or territory (columns 4-6) and 0 otherwise. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 17 African countries in the Bantu region, 56 years (1961-2017), and 86 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis, and are adjusted for the low number of clusters using the number of countries. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.24: *IV Robustness: Controlling for Representation in Government Coalition*

	Ethnic conflict over power			Ethnic conflict over territory		
	OLS	2SLS		OLS	2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Linguistic distance	0.057***	0.180***	0.107**	-0.006	-0.029**	-0.034
	(0.019)	(0.051)	(0.046)	(0.004)	(0.012)	(0.021)
Ethnic group is part of coalition	-0.088	-0.022	-0.052	-0.032	-0.044*	-0.041
	(0.068)	(0.065)	(0.073)	(0.019)	(0.022)	(0.026)
Mean of dep. var.	0.052	0.052	0.052	0.008	0.008	0.008
Distance type	w	w	uw	w	w	uw
Country-year fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes
Geographic controls	yes	yes	yes	yes	yes	yes
Climatic controls	yes	yes	yes	yes	yes	yes
Observations	3,628	3,628	3,628	3,628	3,628	3,628

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power or territory and 0 otherwise. *Ethnic group is part of coalition* is equal to 1 if an ethnicity is represented in the government coalition in a certain year and 0 otherwise. For a description of all explanatory variables, refer to notes in Table 3.2. The sample includes 17 African countries in the Bantu region, 56 years (1961-2017), and 86 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis, and are adjusted for the low number of clusters using the number of countries. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.25: *Reduced Form: Splitting Bantu Distance into Two*

	Ethnic conflict over power		Ethnic conflict over territory	
	(1)	(2)	(3)	(4)
Bantu more	0.059	0.063	-0.001	-0.001
	(0.046)	(0.050)	(0.005)	(0.005)
Bantu less	0.056	0.044	-0.002	-0.004
	(0.058)	(0.059)	(0.003)	(0.004)
F statistics (Bantu more - Bantu less=0)	0.00	0.03	0.00	0.13
Prob > F	0.976	0.861	0.977	0.722
Country fixed effects	yes	yes	yes	yes
Year fixed effects	yes	yes	yes	yes
Country-year fixed effects	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes
Ethnicity year trends	yes	yes	yes	yes
Geographic controls	yes	yes	yes	yes
Climatic controls		yes		yes
Observations	3,628	3,628	3,628	3,628
Adjusted R-squared	0.531	0.534	0.329	0.329

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government and 0 otherwise. *Bantu more* and *Bantu less* is a continuous variable that denotes whether the rebel was more or less exposed to the Bantu expansion than the government, respectively. For a description of geographic controls, refer to the notes in Table 3.2. The sample includes 17 African countries in the Bantu region, 57 years (1961-2017), and 86 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis, and are adjusted for the low number of clusters using the number of countries. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Table C.26: *Principal Component Analysis: Loadings and Sampling Adequacy*

	Loadings						
How well or badly would you say the current government is handling the following matters:	Round 1 1999-2001	Round 2 2002-2003	Round 3 2005-2006	Round 4 2008-2009	Round 5 2011-2013	Round 6 2014-2015	Round 7 2016-2017
Managing the economy		0.299	0.324	0.264	0.273	0.290	0.225
Improving living standards of the poor				0.283	0.283	0.303	0.237
Creating jobs	0.580	0.287	0.311	0.272	0.265	0.293	0.239
Keeping prices stable	0.589	0.286	0.305	0.243	0.233	0.277	0.217
Narrowing the income gap	0.562	0.281	0.312	0.269	0.266	0.286	0.232
Reducing crime		0.288	0.291	0.240	0.246	0.258	0.221
Improving basic health services		0.294	0.312	0.262	0.263	0.278	0.243
Addressing educational needs		0.290	0.306	0.241	0.253	0.271	0.241
Improving water and sanitation services		0.216	0.273	0.244	0.243	0.268	0.216
Ensuring enough to eat		0.272	0.317	0.272	0.264	0.284	0.231
Fighting corruption		0.299	0.309	0.250	0.261	0.277	0.227
Reducing conflict		0.287			0.247		
Combating HIV		0.242	0.249	0.190	0.190		
Combating malaria		0.251					
Maintaining roads and bridges				0.242	0.237	0.255	0.205
Providing reliable electric supply				0.246	0.235	0.258	0.203
Protecting rivers and forests				0.243			
Promoting equal rights and opportunities for women				0.224	0.223		0.226
Preventing election violence							0.238
Preventing or resolving violent community conflict							0.243
Countering violence from armed extremists							0.228
Addressing needs of youth							0.247
Protecting rights, promoting opportunities for disabled							0.236
Kaiser-Meyer-Olkin measure of sampling adequacy	0.664	0.880	0.887	0.918	0.920	0.928	0.938
Common variance explained by first principal component	61%	34%	37%	35%	35%	43%	36%
N	14,334	21,906	25,333	27,515	51,150	52,536	18,004

Notes: The table illustrates the loadings of each variable from the Afrobarometer survey used for the principal component analysis. Each variable is binary and is equal to 1 if the respondent thinks the government is performing badly or fairly badly in handling each policy goal. Sample sizes denote the full set of respondent answering each question. The final sample in Table 3.9 includes only individuals that were successfully merged with the conflict dataset.

Bibliography

Acemoglu, D., Johnson, S. and Robinson, J. A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, **91** (5), 1369–1401.

Afrobarometer Data (1999-2017).

Agüero, J. M. and Frisancho, V. (2017). Misreporting in Sensitive Health Behaviors and its Impact on Treatment Effects: An Application to Intimate Partner Violence. *Working Paper*.

Aizer, A. (2010). The Gender Wage Gap and Domestic Violence. *American Economic Review*, **100** (4), 1847–59.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and Identity. *The Quarterly Journal of Economics*, **115** (3), 715–753.

Alesina, A., Brioschi, B. and La Ferrara, E. (2021). Violence Against Women: A Cross-cultural Analysis for Africa. *Economica*, **88** (349), 70–104.

—, Giuliano, P. and Nunn, N. (2013). On the Origins of Gender Roles: Women and the Plough. *The Quarterly Journal of Economics*, **128** (2), 469–530.

— and La Ferrara, E. (2005). Ethnic Diversity and Economic Performance. *Journal of Economic Literature*, **43** (3), 762–800.

Allansson, M., Melander, E. and Themnér, L. (2017). Organized Violence, 1989–2016. *Journal of Peace Research*, **54** (4), 574–587.

Anderberg, D. and Rainer, H. (2013). Economic Abuse: A Theory of Intrahousehold Sabotage. *Journal of Public Economics*, **97**, 282 – 295.

—, —, Wadsworth, J. and Wilson, T. (2016). Unemployment and Domestic Violence: Theory and Evidence. *The Economic Journal*, **126** (597), 1947–1979.

Anderson, S. (2018). Legal Origins and Female HIV. *American Economic Review*, **108** (6), 1407–39.

Angelucci, M. (2008). Love on the Rocks: Domestic Violence and Alcohol Abuse in Rural Mexico. *The BE Journal of Economic Analysis & Policy*, **8** (1).

Arbatli, C. E., Ashraf, Q. H. and Galor, O. (2013). The Nature of Civil Conflict. *Available at SSRN 2364127*.

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

Ashraf, Q. and Galor, O. (2013). Genetic Diversity and the Origins of Cultural Fragmentation. *American Economic Review*, **103** (3), 528–33.

Ba, I. and Bhopal, R. (2017). Physical, Mental and Social Consequences in Civilians who have Experienced War-related Sexual Violence: a Systematic Review (1981–2014). *Public Health*, **142**, 121–135.

Bahgat, K., Nordås, R. and Østby, G. (2016). Geocoded SVAC Dataset (GEO-SVAC) version 1.0. Peace Research Institute Oslo (PRIO).

Becker, A. (2018). On the Economic Origins of Constraints on Women's Sexuality.

—, Enke, B. and Falk, A. (2020). Ancient Origins of the Global Variation in Economic Preferences. *AEA Papers and Proceedings*, **110**, 319–23.

Becker, G. S. (1957). *The Economics of Discrimination*. Chicago: University of Chicago Press.

Bertocchi, G. and Canova, F. (2002). Did Colonization Matter for Growth?: An Empirical Exploration into the Historical Causes of Africa's Underdevelopment. *European Economic Review*, **46** (10), 1851–1871.

Besley, T. and Reynal-Querol, M. (2014). The Legacy of Historical Conflict: Evidence from Africa. *American Political Science Review*, pp. 319–336.

Bhalotra, S. R., Kambhampati, U. S., Rawlings, S. and Siddique, Z. (2018). Intimate Partner Violence and the Business Cycle. *Working Paper*.

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

Boserup, E. (1970). *Woman's Role in Economic Development*. Earthscan.

Bostoen, K., Clist, B., Doumenge, C., Grollemund, R., Hombert, J.-M., Muluwa, J. K., Maley, J., Blench, R., Di Carlo, P., Good, J. *et al.* (2015). Middle to Late Holocene Paleoclimatic Change and the early Bantu Expansion in the Rain Fforests of Western Central Africa. *Current Anthropology*, **56** (3), 367–368.

—, Grollemund, R. and Muluwa, J. K. (2013). Climate-induced Vegetation Dynamics and the Bantu Expansion: Evidence from Bantu Names for Pioneer Trees (Elaeis guineensis, Canarium schweinfurthii, and Musanga cecropioides). *Comptes Rendus Geoscience*, **345** (7-8), 336–349.

Bremer, S. A. (2000). Who Fights Whom, When, Where, and Why? In J. A. Vasquez (ed.), *What do we Know about War*, Rowman & Littlefield Publishers, pp. 23–36.

Calonico, S., Cattaneo, M. D., Farrell, M. H. and Titiunik, R. (2019). Regression Discontinuity Designs Using Covariates. *The Review of Economics and Statistics*, **101** (3), 442–451.

Campante, F. R., Do, Q.-A. and Guimaraes, B. (2019). Capital Cities, Conflict, and Misgovernance. *American Economic Journal: Applied Economics*, **11** (3), 298–337.

Caselli, F. and Coleman, W. J. (2013). On the Theory of Ethnic Conflict. *Journal of the European Economic Association*, **11** (suppl_1), 161–192.

Cavanaugh, M. M. (2012). Theories of Violence: Social Science Perspectives. *Journal of Human Behavior in the Social Environment*, **22** (5), 607–618.

Cederman, L.-E., Min, B. and Wimmer, A. (2012). Ethnic Armed Conflict dataset.

Cogneau, D. and Moradi, A. (2014). Borders that Divide: Education and Religion in Ghana and Togo since Colonial Times. *Journal of Economic History*, **74** (3), 694–729.

Cohen, D. K. (2013). Explaining Rape During Civil War: Cross-National Evidence (1980–2009). *American Political Science Review*, **107** (3), 461–477.

— and Nordås, R. (2014). Sexual Violence in Armed Conflict: Introducing the SVAC Dataset, 1989–2009. *Journal of Peace Research*, **51** (3), 418–428.

Cools, S. and Kotsadam, A. (2017). Resources and Intimate Partner Violence in Sub-Saharan Africa. *World Development*, **95**, 211–230.

Croicu, M. and Sundberg, R. (2017). UCDP GED Codebook version 18.1. Department of Peace and Conflict Research, Uppsala University.

Dell, M. (2010). The Persistent Effects of Peru's Mining Mita. *Econometrica*, **78** (6), 1863–1903.

—, Lane, N. and Querubin, P. (2018). The Historical State, Local Collective Action, and Economic Development in Vietnam. *Econometrica*, **86** (6), 2083–2121.

Depetris-Chauvin, E. and Özak, Ö. (2020). Borderline Disorder:(De Facto) Historical Ethnic Borders and Contemporary Conflict in Africa. *Working Paper*.

Desmet, K., Ortuño-Ortín, I. and Wacziarg, R. (2012). The Political Economy of Linguistic Cleavages. *Journal of Development Economics*, **97** (2), 322–338.

—, Weber, S. and Ortuño-Ortín, I. (2009). Linguistic Diversity and Redistribution. *Journal of the European Economic Association*, **7** (6), 1291–1318.

Diamond, J. (1997). Guns, Germs, and Steel: The Fates of Human Societies.

Dickens, A. (2018). Ethnolinguistic Favoritism in African Politics. *American Economic Journal: Applied Economics*, **10** (3), 370–402.

— (2020). Understanding Ethnic Differences: The Roles of Geography and Trade.

Duflo, E. (2012). Women Empowerment and Economic Development. *Journal of Economic Literature*, **50** (4), 1051–1079.

Dupraz, Y. (2019). French and British Colonial Legacies in Education: Evidence from the Partition of Cameroon. *The Journal of Economic History*, **79** (3), 628–668.

Erten, B. and Keskin, P. (2018). For Better or for Worse?: Education and the Prevalence of Domestic Violence in Turkey. *American Economic Journal: Applied Economics*, **10** (1), 64–105.

Esteban, J., Mayoral, L. and Ray, D. (2012). Ethnicity and Conflict: An Empirical Study. *American Economic Review*, **102** (4), 1310–42.

— and Ray, D. (2011). Linking Conflict to Inequality and Polarization. *American Economic Review*, **101** (4), 1345–74.

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

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

Fenske, J. (2015). African Polygamy: Past and Present. *Journal of Development Economics*, **117**, 58–73.

Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas. *International Journal of Climatology*, **37** (12), 4302–4315.

Fonchingong, C. C. and Fonjong, L. N. (2003). The Concept of Self-reliance in Community Development Initiatives in the Cameroon Grassfields. *Nordic Journal of African Studies*, **12** (2), 196–219.

Friedl, E. (1978). Society and Sex Roles. *Human Nature*, **1** (4), 68–75.

Galor, O. and Özak, Ö. (2016). The Agricultural Origins of Time Preference. *American Economic Review*, **106** (10), 3064–3103.

—, Özak, Ö. and Sarid, A. (2018). Geographical Roots of the Coevolution of Cultural and Linguistic Traits. *Available at SSRN 3284239*.

Gelman, A. and Imbens, G. (2019). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics*, **37** (3), 447–456.

Gifford, P. and Smith, A. (1967). *Britain and Germany in Africa: Imperial Rivalry and Colonial Rule*. New Haven: Yale University Press.

Girardin, L., Hunziker, P., Cederman, L.-E., Bormann, N.-C. and Vogt, M. (2015). Growup-Geographical Research on War, Unified Platform. *ETH Zurich. http://growup. ethz. ch*.

Giuliano, P. (2018). *Gender: A Historical Perspective. In: The Oxford Handbook of Women and the Economy*. Edited by Susan L. Averett, Laura M. Argys, and Saul D. Hoffman. New York: Oxford University Press.

Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M. and Strand, H. (2002). Armed Conflict 1946-2001: A New Dataset. *Journal of Peace Research*, **39** (5), 615–637.

Gneezy, U., Leonard, K. L. and List, J. A. (2009). Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society. *Econometrica*, **77** (5), 1637–1664.

Gomes, J. F. (2020). The Health Costs of Ethnic Distance: Evidence from Sub-Saharan Africa. *Journal of Economic Growth*, **25** (2), 195–226.

Goodman-Bacon, A. (2018). Difference-in-differences with Variation in Treatment Timing. *Working Paper*.

Gouvernement Francais (1922). Rapport Annuel sur l'Administration sous Mandat des Territories du Cameroun pour l'Annee 1922.

Gouvernement Francais (1923). Rapport Annuel sur l'Administration sous Mandat des Territories du Cameroun pour l'Annee 1923.

Gouvernement Francais (1927). Rapport Annuel sur l'Administration sous Mandat des Territories du Cameroun pour l'Annee 1927.

Government of the United Kingdom (1958a). Annual Report of the Cameroon Development Corporation for the year 1958.

Government of the United Kingdom (1958b). The Cameroons under United Kingdom Administration: Report for the year 1958.

Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C. and Pagel, M. (2015). Bantu Expansion Shows that Habitat Alters the Route and Pace of Human Dispersals. *Proceedings of the National Academy of Sciences*, **112** (43), 13296–13301.

Grosjean, P. (2011). The Weight of History on European Cultural Integration: a Gravity Approach. *American Economic Review*, **101** (3), 504–08.

Guariso, A. and Rogall, T. (2017). Rainfall Inequality, Political Power, and Ethnic Conflict in Africa. *LICOS Discussion Paper Series*.

Guarnieri, E. and Tur-Prats, A. (2020). Cultural Distance and Conflict-Related Sexual Violence.

Guiso, L., Sapienza, P. and Zingales, L. (2016). Long-term Persistence. *Journal of the European Economic Association*, **14** (6), 1401–1436.

Guryan, J. (2001). Desegregation and Black Dropout Rates. *National Bureau of Economic Research*.

Hallward-Driemeier, M., Hasan, T., Kamangu, J., Lobti, E. and Blackden, M. (2012). Women's Legal and Economic Empowerment Database for Africa. *Washington, DC: World Bank*.

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

Harbom, L., Melander, E. and Wallensteen, P. (2008). Dyadic Dimensions of Armed Conflict, 1946—2007. *Journal of Peace Research*, **45** (5), 697–710.

Hidrobo, M., Peterman, A. and Heise, L. (2016). The Effect of Cash, Vouchers, and Food Transfers on Intimate Partner Violence: Evidence from a Randomized Experiment in Northern Ecuador. *American Economic Journal: Applied Economics*, **8** (3), 284–303.

Holsoe, S. E. and Lauer, J. J. (1976). Who Are the Kran/Guere and the Gio/Yacouba? Ethnic Identifications Along the Liberia-Ivory Coast Border. *African Studies Review*, **19** (1), 139–150.

Horowitz, D. L. (2000). *Ethnic Groups in Conflict, updated edition with a new preface*. Univ of California Press.

Huillery, E. (2009). History Matters: The Long-Term Impact of Colonial Public Investments in French West Africa. *American Economic Journal: Applied Economics*, **1** (2), 176–215.

Huntington, S. P. (2000). The Clash of Civilizations? In *Culture and Politics*, Springer, pp. 99–118.

Institute for Economics & Peace (2020). Global Peace Index 2020: Measuring Peace in a Complex World. `http://visionofhumanity.org/reports`.

International Criminal Court (2000). Elements of Crimes. PCNICC /2000/1/Add.2.

International Criminal Court (2002). Rome Statute. Document A/CONF.183/9.

International Food Policy Research Institute (IFPRI) (2015). Agro-Ecological Zones for Africa South of the Sahara.

Koenig, M. A., Ahmed, S., Hossain, M. B. and Mozumder, A. K. A. (2003). Women's Status and Domestic Violence in Rural Bangladesh: Individual- and Community-level Effects. *Demography*, **40** (2), 269–288.

Konings, P. (2012). *Gender and Plantation Labour in Africa: the Story of Tea Pluckers' Struggles in Cameroon*. African Books Collective.

Korotayev, A. (2003). Division of Labor by Gender and Postmarital Residence in Cross-cultural Perspective: A Reconsideration. *Cross-Cultural Research*, **37** (4), 335–372.

La Mattina, G. (2017). Civil Conflict, Domestic Violence and Intra-household Bargaining in Post-genocide Rwanda. *Journal of Development Economics*, **124**, 168–198.

— and Shemyakina, O. N. (2017). Domestic Violence and Childhood Exposure to Armed Conflict: Attitudes and Experiences. *Working Paper*.

La Porta, R., Lopez-de Silanes, F., Shleifer, A. and Vishny, R. (1999). The Quality of Government. *The Journal of Law, Economics, and Organization*, **15** (1), 222–279.

—, —, — and Vishny, R. W. (1997). Legal Determinants of External Finance. *The Journal of Finance*, **52** (3), 1131–1150.

—, —, — and — (1998). Law and Finance. *Journal of Political Economy*, **106** (6), 1113–1155.

Le Vine, V. T. (1964). *The Cameroons, from Mandate to Independence*. University of California Press.

Lee, A. and Schultz, K. A. (2012). Comparing British and French Colonial Legacies: A Discontinuity Analysis of Cameroon. *Quarterly Journal of Political Science*, **7**, 1–46.

Lee, D. L., McCrary, J., Moreira, M. J. and Porter, J. (2020). Valid t-ratio Inference for IV. *arXiv preprint arXiv:2010.05058*.

Leonard, K. E. (2005). Alcohol and Intimate Partner Violence: When can we say that Heavy Drinking is a Contributing Cause of Violence? *Addiction*, **100** (4), 422–425.

Lesthaeghe, R. J. (1989). *Reproduction and Social Organization in sub-Saharan Africa*, vol. 4. Univ of California Press.

Lewis, M. P., Simons, G. F. and Fennig, C. D. (2020). Ethnologue: Languages of the World. *Dallas, Texas: SIL International. Online version: http://www.ethnologue.com*.

Louis, W. R. (1967). *Great Britain and Germany's Lost Colonies, 1914-1919*. Oxford: Clarendon.

— and Gifford, P. (1971). *France and Britain in Africa: Imperial Rivalry and Colonial Rule*. Yale University Press.

Lowes, S. (2017). Matrilineal Kinship and Spousal Cooperation: Evidence from the Matrilineal Belt. *Unpublished manuscript. URL: https://scholar. harvard. edu/files/slowes/files/lowes_matrilineal. pdf*.

— and Montero, E. (2020). *Concessions, Violence, and Indirect Rule: Evidence from the Congo Free State*. Working Paper 27893, National Bureau of Economic Research.

Martin, M. K. and Voorhies, B. (1975). *Female of the Species*. Columbia University Press.

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

Meger, S. (2016). *Rape Loot Pillage: The Political Economy of Sexual Violence in Armed Conflict*. Oxford University Press.

Michalopoulos, S. (2012). The Origins of Ethnolinguistic Diversity. *The American economic review*, **102** (4), 1508–1539.

— and Papaioannou, E. (2013). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica*, **81** (1), 113–152.

— and — (2014). National Institutions and Subnational Development in Africa. *The Quarterly Journal of Economics*, **129** (1), 151–213.

— and — (2016). The Long-run Effects of the Scramble for Africa. *The American Economic Review*, **106** (7), 1802–1848.

—, Putterman, L. and Weil, D. N. (2018). The Influence of Ancestral Lifeways on Individual Economic Outcomes in Sub-Saharan Africa. *Journal of the European Economic Association*, **17** (4), 1186–1231.

Mokyr, J. (2016). *A Culture of Growth: the Origins of the Modern Economy*. Princeton University Press.

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

— and — (2017). Ethnic Diversity and Growth: Revisiting the Evidence. *Review of Economics and Statistics*, pp. 1–43.

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

Müller-Crepon, C., PenglL, Y. and Bormann, N.-C. (Forthcoming). Linking Ethnic Data from Africa. *Journal of Peace Research*.

Muñoz-Encinar, L. (2019). Unearthing Gendered Repression: An Analysis of the Violence Suffered by Women during the Civil War and Franco's Dictatorship in Southwestern Spain. *World Archaeology*, **51** (5), 759–777.

Munshi, K. and Rosenzweig, M. (2006). Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy. *American Economic Review*, **96** (4), 1225–1252.

Murdock, G. P. (1967). Ethnographic Atlas: A Summary. *Ethnology*, **6** (2), 109–236.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, pp. 1417–1426.

Nowrojee, B. (1996). *Shattered Lives: Sexual Violence during the Rwandan Genocide and its Aftermath*, vol. 3169. Human Rights Watch.

Nunn, N. (2008). The Long Term Effects of Africa's Slave Trades. *Quarterly Journal of Economics*, **123** (1), 139–176.

— (2010). Religious Conversion in Colonial Africa. *American Economic Review Papers and Proceedings*, **100** (2), 147–152.

— (2014a). *Gender and Missionary Influence in Colonial Africa*, New York: Cambridge University Press, pp. 489–512.

— (2014b). Historical Development. In P. Aghion and S. Durlauf (eds.), *Handbook of Economic Growth*, vol. 2, North-Holland, pp. 347–402.

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

Palermo, T., Bleck, J. and Peterman, A. (2013). Tip of the Iceberg: Reporting and Gender-based Violence in Developing Countries. *American Journal of Epidemiology*, **179** (5), 602–612.

Pfeifer, G. and Reutter, M. (2020). The Morning After: Prescription-Free Access to Emergency Contraceptive Pills. *Working Paper*.

Preston, P. (2012). *The Spanish Holocaust: Inquisition and Extermination in Twentieth-century Spain*. WW Norton & Company.

Putterman, L. and Weil, D. N. (2010). Post-1500 Population Flows and the Long-run Determinants of Economic Growth and Inequality. *The Quarterly Journal of Economics*, **125** (4), 1627–1682.

Qian, N. (2008). Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance. *Quarterly Journal of Economics*, **123** (3), 1251–1285.

Robbeets, M. and Savelyev, A. (2017). *Language Dispersal Beyond Farming*. John Benjamins Publishing Company.

Sanday, P. R. (1973). Toward a Theory of the Status of Women. *American Anthropologist*, **75** (5), 1682–1700.

— (1981a). *Female Power and Male Dominance: On the Origins of Sexual Inequality*. Cambridge University Press.

— (1981b). The Socio-cultural Context of Rape: A Cross-cultural Study. *Journal of Social Issues*, **37** (4), 5–27.

Schneider, D. M. and Gough, K. (1961). *Matrilineal Kinship*. Univ of California Press.

Serafinelli, M. and Tabellini, G. (2017). Creativity over Time and Space. *Available at SSRN 3070203*.

SIGI (2019). *The Social Institutions and Gender Index (SIGI) Cameroon Country Profile*. Tech. rep.

Skjelsbaek, I. (2001). Sexual Violence and War: Mapping Out a Complex Relationship. *European Journal of International Relations*, **7** (2), 211–237.

Smith, M. G. (1965). *The Plural Society in the British West Indies*. Univ of California Press.

Spolaore, E. and Wacziarg, R. (2009). The Diffusion of Development. *The Quarterly journal of economics*, **124** (2), 469–529.

— and — (2016a). Ancestry, Language and Culture. In *The Palgrave Handbook of Economics and Language*, Springer, pp. 174–211.

— and — (2016b). War and Relatedness. *Review of Economics and Statistics*, **98** (5), 925–939.

— and — (2017). *The Political Economy of Heterogeneity and Conflict*. Tech. rep., National Bureau of Economic Research.

Straus, M. A. (1979). Measuring Intrafamily Conflict and Violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and the Family*, pp. 75–88.

—, Hamby, S. L., Boney-McCoy, S. and Sugarman, D. B. (1996). The Revised Conflict Tactics Scales (CTS2) Development and Preliminary Psychometric Data. *Journal of Family Issues*, **17** (3), 283–316.

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

Symons, D. (1979). The Evolution of Human Sexuality.

Taylor, C. C. (1999). A Gendered Genocide: Tutsi Women and Hutu Extremists in the 1994 Rwanda Genocide. *PoLAR*, **22**, 42.

Tertilt, M. (2005). Polygyny, Fertility, and Savings. *Journal of Political Economy*, **113** (6), 1341–1371.

Teso, E. (2018). The Long-term Effect of Demographic Shocks on the Evolution of Gender Roles: Evidence from the Transatlantic Slave Trade. *Journal of the European Economic Association*, **17** (2), 497–534.

Thornhill, R. and Palmer, C. T. (2001). *A Natural History of Rape: Biological Bases of Sexual Coercion*. MIT press.

Tur-Prats, A. (2017). Unemployment and Intimate-Partner Violence: A Gender Identity Approach. *UPF Discussion Paper No. 1564*.

— (2019). Family Types and Intimate Partner Violence: A Historical Perspective. *Review of Economics and Statistics*, **101** (5), 878–891.

United Nations Office on Drugs and Crime (2018). Global Study on Homicide 2018.

United Nations Trusteeship Council (1949). Report of the Trusteeship Council covering its Fourth and Fifth Sessions 6 August 1948-22 July 1949.

United Nations Trusteeship Council (1950a). Notes to the Eightieth meeing of the Trusteeship Council: 3 April 1950.

United Nations Trusteeship Council (1950b). Report of the First United Nations Visiting Mission to the Trust Territory of the Cameroons under French Administration.

United Nations Trusteeship Council (1954). United Nations Visiting Mission to Trust Territories in West Africa, 1952 : Report on the Cameroons under French Administration, Together with Related Documents.

United Nations Trusteeship Council (1956). Report of the Trusteeship Council covering the Period from 23 July 1955 to 14 August 1956.

United Nations Trusteeship Council (1958). United Nations Visiting Mission to Trust Territories in West Africa, 1958 : report on the Cameroons under French Administration, Together with Related Documents.

United States Department of State (2017). *Cameroon 2017 Human Rights Report*. Tech. rep.

Vogt, M., Bormann, N.-C., Rüegger, S., Cederman, L.-E., Hunziker, P. and Girardin, L. (2015). Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family. *Journal of Conflict Resolution*, **59** (7), 1327–1342.

Voigtländer, N. and Voth, H.-J. (2012). Persecution Perpetuated: the Medieval Origins of anti-Semitic Violence in Nazi Germany. *The Quarterly Journal of Economics*, **127** (3), 1339–1392.

Wimmer, A., Cederman, L.-E. and Min, B. (2009). Ethnic Politics and Armed Conflict: A Configurational Analysis of a New Global Data Set. *American Sociological Review*, **74** (2), 316–337.

Wood, E. J. (2009). Armed Groups and Sexual Violence: When is Wartime Rape Rare? *Politics & Society*, **37** (1), 131–161.

World Health Organization (2001). Ethical and Safety Recommendations for Research on Domestic Violence Against Women. *Geneva, Switzerland: World Health Organization*.

Wucherpfennig, J., Metternich, N. W., Cederman, L.-E. and Gleditsch, K. S. (2012). Ethnicity, the State, and the Duration of Civil War. *World Politics*, **64** (1), 79–115.

—, Weidmann, N. B., Girardin, L., Cederman, L.-E. and Wimmer, A. (2011). Politically Relevant Ethnic Groups across Space and Time: Introducing the GeoEPR Dataset. *Conflict Management and Peace Science*, **28** (5), 423–437.

Yearwood, P. J. (1993). "In Casual Way with a Blue Pencil": British Policy and the Partition of Kamerun, 1914–1919. *Canadian Journal of African Studies/La Revue canadienne des études africaines*, **27** (2), 218–244.