Nation-Building in Sub-Saharan Africa and Civil Conflict: Theory and Evidence from *Boko Haram* and *Tuareg* Insurgencies*

Maxim Ananyev† Michael Poyker‡

December 2016

Click here for the latest version

Abstract

Why is civil conflict so costly for development? We argue that civil conflict undermines the legitimacy of the nation-state and empowers traditional sources of authority. In particular, we demonstrate, using recent instances of an insurgency in West Africa, that civil conflict erodes national identities, replacing them by ethnic identities. Based on the existing historical, anthropological, and ethnographic evidence, we model the choice of loyalty (national or ethnic) as a coordination game with strategic complementarities ("global game"). This model allows us to show how the instances of civil conflict can break up that coordination and impede nation-building. We perform several estimation strategies (including difference-in-difference and instrumental variables) to quantify the effect of civil conflict of national identity in three nations: Burkina Faso, Mali, and Nigeria. The identification of the effect comes from using pre-independence data on the location of ethnic homelands of rebellious groups of *Tuareg* (in case of Burkina Faso and Mali) and *Hausa/Fulani* (in case of Nigeria). Our key assumption is that the location of those groups in colonial times is independent of the "potential outcome": potential changes in national identity between years 2010 and 2012. We explore the plausibility of this assumption using pre-treatment trends, placebo tests, and robustness checks. We also find that our estimates are resilient to the violation of exclusion restrictions (even the violation is large as our most important individual-level predictor of national identity does not revert our findings). Our theory and evidence contribute to the study of state formation and state capacity by exploring the roots of people's self-identification with a state.

JEL Codes: D19, J15, O12, Z13

Keywords: National Identity, Insurgency, *Tuaregs*, *Boko Haram*

PRELIMINARY AND INCOMPLETE. COMMENTS ARE WELCOME

*The authors are grateful to Ruth Carlitz, Daniel Posner, Vera Mironova and conference participants at ASREC 2016, MPSA 2016 and UC GEM-BPP Research Workshop 2016 for excellent suggestions. All errors are ours. Correspondence at: UCLA Anderson School of Management, 110 Westwood Plaza, C502 Entrepreneurs Hall, Los Angeles, CA 90095-1481.

†Department of Political Science, UCLA; e-mail address: maksim.ananjev@gmail.com

‡UCLA Anderson School of Management; e-mail address: mikhail.poyker.1@anderson.ucla.edu

1 Introduction

One of the most remarkable and often under-appreciated achievements of African post-colonial development is the construction of national identities. While many of the states in Africa have arbitrary borders drawn by European colonial powers, preoccupied with their own rivalry, in the course of the post-colonial development those *imagined* borders, became a *reality* in two important respects. First, the African political map proved remarkably resilient with comparatively small number of conflicts between states. Second, according to the Afrobarometer survey, most of the people in Sub-Saharan Africa, view their national identity at least as important as their ethno-linguistic group.

How this construction of nations become possible and what are its main impediments? In this paper, we offer a formal model of nation-building that puts information and coordination in the heart of this process. In the spirit of the definition of state as an organization that controls violence, first coined by Leo Trotsky and then famously borrowed by Max Weber, we show that civil violence can break up the successful coordination of identities.

The inspiration for our model also comes from the idea of "two publics" by Peter P. Ekeh. Ekeh [1975] asserts that African nations are caught in the duality of loyalties: the first loyalty is to the "primordial public" (the word "primordial" here is used by Ekeh for expressive convenience, with the full appreciation of socially constructed nature of many of Sub-Saharan ethno-linguistic groups), and the second one is to the "civic public". This conflict of loyalties shapes many of the aspects of political and economic development in Africa. Building on this insight, we outline the formal logic of the choice between the two loyalties, and demonstrate how incapacity of the state to control violence shifts the balance away from the civic public.

We believe that our study has broad theoretical and practical implications. Nation-state is one of the most important inventions of modern era. Much of people's political, economic, and social life is organized around the existence of geographically defined, "sovereign" entities. Construction of such sovereign entities – nation-states – is important for economic development. Without the capable state, it is nearly impossible to ensure the provision public goods and services, enforcement of contracts, and national security (Dincecco and Prado [2012], Acemoglu et al. [2014]). A number of studies have looked at the problem of nation-building from the perspective of tangible tasks that a state needs to perform, like education, administration, police, and taxation (Geddes [1994], Besley and Persson [2010], Soifer [2015]). Those studies help us understand the challenges the governments of weak states face when they try to perform functions that are normally seen as the state's responsibility.

One of the understudied aspects of state-building is a social construction of nation, an "imagined community" that the citizens of a country see themselves belonging to. Sociologists, political theorists, and historians have been contributing to this question (Anderson [2006], Gellner and Breuilly [2008], Hobsbawm [2012], Mann [2012], Miguel [2004], Robinson [2014]). Many studies have explored the construction of identities using data from the developing nations.(Laitin [1998], Posner [2004], Miguel [2004], Eifert et al. [2010], Robinson [2014]). Our study contributes to this effort.

In the literature, the study of the impact of the warfare on national identities is often confined to the effects of *international* warfare.[1] For example, Dell and Querubin [2016] show that territories affected by US bombardments during the Vietnam War were experiencing a rise in the communist insurgency. Similarly, civil wars are believed sometimes to be nation-defining events, out of which a nation emerges. For example, scholars believe that American Civil War "forged" American identity (Huntington [2004], Neely [2011], Vorenberg [2013]), or that Spanish Civil War (and subsequent reconciliation) contribute to Spanish identity (Aguilar Fernández and Humlebaek [2002]).

In this paper, we attempt to advance the study of national identity in two respects: first, we offer a formal model, where we theorize national identity as a coordination problem. In our setup, citizens would like to become a part of nation if the other citizens make the same choice. Because the viability of national identity is anchored in the existence of a well-functioning state, the threats to state institutions, break up that coordination, and thus erode national identity. Secondly, we test the main prediction of the model – negative association between the insurgency and national identity – using the survey data from the regions in West Africa that have been recently exposed to the ethnic or religious insurgency.

In our model, by using a global games approach, exemplified in Morris and Shin [1998, 2004] we assume, that each citizen of a multi-ethnic state has a choice: to keep their ethnic identity (status quo), or embrace national identity with potential benefits in case enough people will decide to support the national identity. Each citizen has their own private estimate of the benefits from accepting national identity and observe a public signal about the strength of the state. If a state becomes observably weaker, it decreases the expected benefit from accepting the national identity.

Motivated by the predictions of the theoretical model, we employ difference-in-difference approach to estimate an effect of insurgency on the probability of accepting national identity by the respondents of Afrobarometer surveys. We use survey data from the places affected by two recent insurgencies in sub-Saharan Africa. The first one is the *Tuareg* rebellion caused by the fact, that *Tuaregs* were first hired by Muamar al-Gaddafi when the Lybian Civil War started, and then returned to their homelands with the weapons when the Libyan regime had been defeated. The second one is a *Boko Haram* insurgency in the northern Nigeria, that started after the death of one of the local Muslim leaders in 2009. We use Afrobarometer surveys that were completed before and after insurgency, and show that in the regions affected by insurgency probability of accepting national identity decreased. An important part of our identification strategy is avoiding a post-treatment bias by assigning regions to "treatment" and "control" groups based on the information on the ethnic composition of those regions (in particular, we use the information from *Atlas Narodov Mira* about ethnic homelands of *Tuaregs* and ethnic groups involved in *Boko Haram*). Finally, we employ IV estimation, by using historical homelands of ethnic groups involved in insurgency and temporal variation in rebellion as instruments for the insurgency. For the reasons of data availability, we analyze three countries: Burkina Faso, Mali, and Nigeria. In our preferred specification, the one standard deviation in the

[1] For those interested in researching the topic beyond the material covered in this paper, possibly the best reference to consult is Bauer et al. [2016].

exposure to an insurgency results into -0.43 standard deviation of national self-identification on the regional level.

The paper is organized as follows. The national identity choice model is introduced in Section 2. Background information about *Tuaregs* and *Boko Haram* is presented in Section 3. Section 4 describes the data. Section 5 explains identification strategy. Section 6 contains main empirical results. Robustness and sensitivity checks are presented in Section 7. Section 8 contains concluding remarks.

2 Model: National Identity as a Coordination Problem

We use the "global games" approach, proposed in Carlsson and Van Damme [1993] and popularized in Morris and Shin [1998, 2004], to model the choice of identity by the residents of a country exposed to a civil conflict. We view the choice of identity as a coordination problem. People of different ethnicities can either choose a national identity or keep their ethnic identity as their primary identity. Here, we do not assume that ethnic identity is primordial in any sense, just that the construction of ethnic identities had preceded the nation-building efforts by post-colonial governments. . The payoffs from having a national identity depend on how many other people also choose national identity, and are potentially larger than the payoffs from ethnic identity. The payoffs from national identity depend on the strength of the state. If a state is not able to contain violence, then the payoff goes down. This is why the information about insurgent activities has a twofold effect: it directly decreases the utility of associating oneself with the state, and it also inhibits the coordination on a national identity, because it influences higher order beliefs about whether the other people in a society are likely to prefer national identity.

The following subsections formalize this intuition.

Setup

Consider an area with a continuum of agents-residents of a country (i) of the measure one, uniformly distributed over $[0, 1]$. We assume that all agents are homogeneous, but might belong to different ethnic groups. Agents face a binary choice of whether to accept a national identity or to retain an ethnic identity.

Resident i's payoff as follows:

$$u_i = \begin{cases} 0 & \text{if the agent keeps ethnic identity} \\ \theta + \tau A & \text{if the agent accepts national identity} \end{cases} \tag{2.1}$$

where parameter θ captures the expected net benefit from strong national state. Here we assume that ethnic identity is a status quo: people have innate preferences for being a part of their "tribe", as their parents and grandparents. Importantly, we allow "group size benefit" for those who decide to associate themselves with the nation: their payoff depends on $A \in [0, 1]$ the total share of agents deciding to associate themselves with the country's nation. An example, of such benefit is an access to credit and urban jobs. It

has been documented that in many African nations, ethnicity serves as a vehicle for patronage dispension. In case of Zambia, Posner [2005] describes a tendency of ethnical favoritism in hiring, succintly put in a Nyanja term "wako ni wako" ("what yours is yours"). Parameter τ captures the the magnitude of this benefit from the size.

The agents have a diffuse prior distribution of $\theta \in \mathbb{R}$. We assume, that agents live in two regions. Share r of the citizens live in the region where insurgency occurs, and share of $(1-r)$ citizens live in a region of the country without the insurgency. All citizens receive common signal $p_j = \theta + \varepsilon_j$, where $\varepsilon_I \sim \mathcal{N}(0, \alpha_I)$ is the noise in the region with insurgency, and $\varepsilon_{NI} \sim \mathcal{N}(0, \alpha_{NI})$ is the noise in the region without insurgency. It can be interpreted as an imperfect common knowledge about the stability of the government. Here we assume that citizens of the region with the insurgency will have higher precision about the insurgency ($\alpha_I < \alpha_{NI}$). Citizens of the region without the insurgency assume that citizens of other region have the similar public signal that they have received. At the same time, citizens of the affected region know that their signal is more precise and observe all information that individuals in non-affected region see (p^{NI}, α_{NI}).

In addition, each citizen receives an independent private signal $x_i = \theta + \xi_i$, where $\xi_i \sim \mathcal{N}(0, \beta)$, is noise. It represents a personal assessment of the benefits from national identity (in vein of Carvalho [2010]). People update their private signal with the public one: $\bar{\theta_i^j} = \theta | p^j, x_i \sim \mathcal{N}\left(\frac{\beta x_i + \alpha_j p}{\alpha_j + \beta}; \quad \frac{\alpha_j \beta}{\alpha_j + \beta}\right), j \in \{I, NI\}$.

Analysis

We seek equilibrium in threshold strategies. Agents have two strategies:

$$a_i(\bar{\theta_i^j}) = \begin{cases} \text{keep ethnic identity} & \text{if } \bar{\theta_i^j} \leq \kappa^{j*} \\ \text{accept national identity} & \text{if } \bar{\theta_i^j} > \kappa^{j*} \end{cases} \tag{2.2}$$

That is agents accept national identity if their beliefs about future benefits from sharing national identity is high, i.e. $\bar{\theta_i}$ is above some threshold κ^*. There is a unique equilibrium cutoff κ for the citizens defined by the expectation of future benefits from national state that makes an individual indifferent regarding the choice of accepting national or tribal identity.

The equilibrium participation threshold κ is the solution to the equilibrium condition:

$$\kappa^{NI} + \tau \Phi\left(\sqrt{\gamma^{NI}}(p^{NI} - \kappa^{NI})\right) = 0, \tag{2.3}$$

$$\kappa^{I} + \tau \left(r\Phi\left(\sqrt{\gamma^{I}}(p^{I} - \kappa^{I})\right) - \frac{(1-r)\kappa^{NI}}{\tau}\right) = 0 \tag{2.4}$$

where $\gamma^j = \frac{\beta\left(\alpha^j + \beta\right)}{\alpha^{j2}(\beta + 2\alpha^j)}, \quad j \in \{I, NI\}$. As shown in Morris and Shin [1998, 2004] the equilibrium is unique if regularity conditions ($\tau^2 \gamma < 2\pi$) hold (see the proof in the Appendix A.1).

Comparative statics

After we have found the equilibrium thresholds κ^* we can derive the equilibrium participation, A:

$$A^I = \Phi\left(\frac{\frac{\beta}{\alpha^I}p^I + \theta - \frac{\beta+\alpha^I}{\alpha^I}\kappa^I}{\beta}\right) \tag{2.5}$$

It is continuous and strictly increasing in p.

We find the probability that an agent will choose a national identity:

$$Prob(a_i(\bar{\theta_i^I}) = 1) = \Phi\left(\frac{\beta}{\sqrt{\beta+\alpha^I}} \cdot x_i + \frac{\alpha^I}{\sqrt{\beta+\alpha^I}} \cdot p^I - \sqrt{\beta+\alpha^I} \cdot \bar{\theta_i^I}\right) \tag{2.6}$$

Proposition 1. *: Share of people who accept national identity is always increasing in signal about about the strength of the state ($\frac{\partial A}{\partial p} > 0$).*

Proof:

$$\frac{\partial A^I}{\partial p^I} = \frac{1}{\alpha^I}\phi\left(\frac{\frac{\beta}{\alpha^I}p^I + \theta - \frac{\beta+\alpha^I}{\alpha^I}\kappa^I}{\beta}\right) > 0 \tag{2.7}$$

Proposition 2. *: Probability of an agent to accept a national identity is always increasing in signal about the strength of the state ($\frac{\partial\left[Prob(a_i(\bar{\theta_i^I})=1)\right]}{\partial p^I} > 0$).*

Proof:

$$\frac{\partial\left[Prob(a_i(\bar{\theta_i^I}) = 1)\right]}{\partial p^I} = \frac{\alpha^I}{\sqrt{\beta+\alpha^I}}\phi\left(\frac{\beta}{\sqrt{\beta+\alpha^I}} \cdot x_i + \frac{\alpha^I}{\sqrt{\beta+\alpha^I}} \cdot p^I - \sqrt{\beta+\alpha^I} \cdot \bar{\theta_i^I}\right) > 0 \tag{2.8}$$

For the purposes of this paper, this model produces a specific testable prediction: the probability of a person identifying with the nation should decrease when the state becomes weaker.

3 *Tuareg* Rebellion and *Boko Haram*: Background Information

This paper uses two recent instances of civil conflict in Sub-Saharan Africa to test the predictions of the model: in particular, the relationship between the national identity and the exposure to civil conflict. This section offers a brief introduction to the substantive background of the cases we consider in our empirical tests.

Figure 3.1 depicts the map of African continent, where the changes in the number of terrorist attacks in particular country are shown in red. We see that countries of Burkina Faso, Mali, and Nigeria are among the

nations that has recently suffered from civil conflicts. However, this aggregate picture masks an important within-country variation. In Burkina Faso, the insurgency happened mostly in the northern provinces of *Odulam*, *Soum*, and *Seno* (Figure 3.2). In Mali, the insurgency happened in northern regions of *Toumbuktu*, *Kidal*, and *Gao*, collectively known as *Azawad* (Figure 3.2). In Nigeria, the insurgency happened mostly in the north-east states of *Borno*, *Yobe*, and *Adamawa* (Figure 3.2).

Why the insurgencies happened in some places but less so – in others? The location and timings of the insurgencies were not random, it dependent several important factors. First, the insurgencies happened in the ethnic homelands of particular groups: *Tuaregs* (in case of Burkina Faso and Mali) and *Hausa/Fulani* (in case of Nigeria). Figure 3.3 shows the areas of *Tuareg* and *Hausa/Fulani* presence based on Soviet *Atlas Narodov Mira*. The timing of some of those rebellions was also not random, it was influences by the abrupt regime change in Libya.

Civil war in Libya wreaked havoc in the Magreb and West Africa. In the time of unrest, colonel al-Gaddafi used his vast financial resources to train, arm, and fund large numbers of *Tuaregs* – semi-nomadic ethnic minority group.[2] When he died, the *Tuaregs* took the guns back to their homelands: regions of Algeria, Mali, Niger and Burkina Faso and attempted to take control of that territory.[3] Some groups of *Tuaregs* went southward in this semi-arid belt of land known as the Sahel.

In Mali, they led a full-fledged rebellion and, for a time, seized the country's northern half. For example, a *Tuareg* group that calls itself *National Movement for the Liberation of Azavad* was making a military assault on the targets not only in Mali but also in Niger and Burkina Faso, and reached as far as Niamey in the South-western Niger.[4]

As a visual example of insurgency, we present a graph (Figure 3.4) of total number of killed and wounded people[5] due to terrorist attacks in Mali, Niger, Nigeria, and Burkina Faso. Because Tuareg migration happened in 2011, for our difference-in-difference estimation (to be described later), we designate the year 2010 as a year "before treatment", and the year 2012 as year "after treatment".

The second example of the effect of insurgency on the national identification comes from the activity of terrorist organization *Boko Haram* in Nigeria (Umar [2011]). This radical extremist Islamic organization, based in northeastern Nigeria, is also active in Chad, Niger[6] and northern Cameroon It was founded in 2002, however terrorist activity started only in 2009. It adheres to the *Salafi jihadism* ideology of Islam, that in turn originated from the *Wahhabi* denomination of Sunni Islam and spread through other Muslim denominations. In addition to high violence, it can be characterized by opposition to the perceived "westernization" of

[2]More information about this fact can be found here `https://www.opendemocracy.net/hugh-brody/gaddafi-and-tuareg-lords-of-desert`, here `http://www.thecuttingedgenews.com/index.php?article=71898&pageid=20&pagename=Security` or here `http://www.nytimes.com/2012/02/06/world/africa/tuaregs-use-qaddafis-arms-for-rebellion-in-mali.html?_r=0`.

[3]*http://www.webcitation.org/6AGfVdHeO*

[4]*http://foreignpolicy.com/2012/04/05/the-mess-in-mali/*

[5]Based on the data by Global Terrorist Database (GTD [2015]).

[6]It conquered three Niger's cities: Bosso, Diffa and N'guigme by January 2015. However, *Boko Haram* didn't have significant effect on Niger during the 2010-2012 period that was used in the regression setting 5.2.

Figure 3.1: Severity of the terrorist attacks in Africa (in changes between 2010 and 2012)

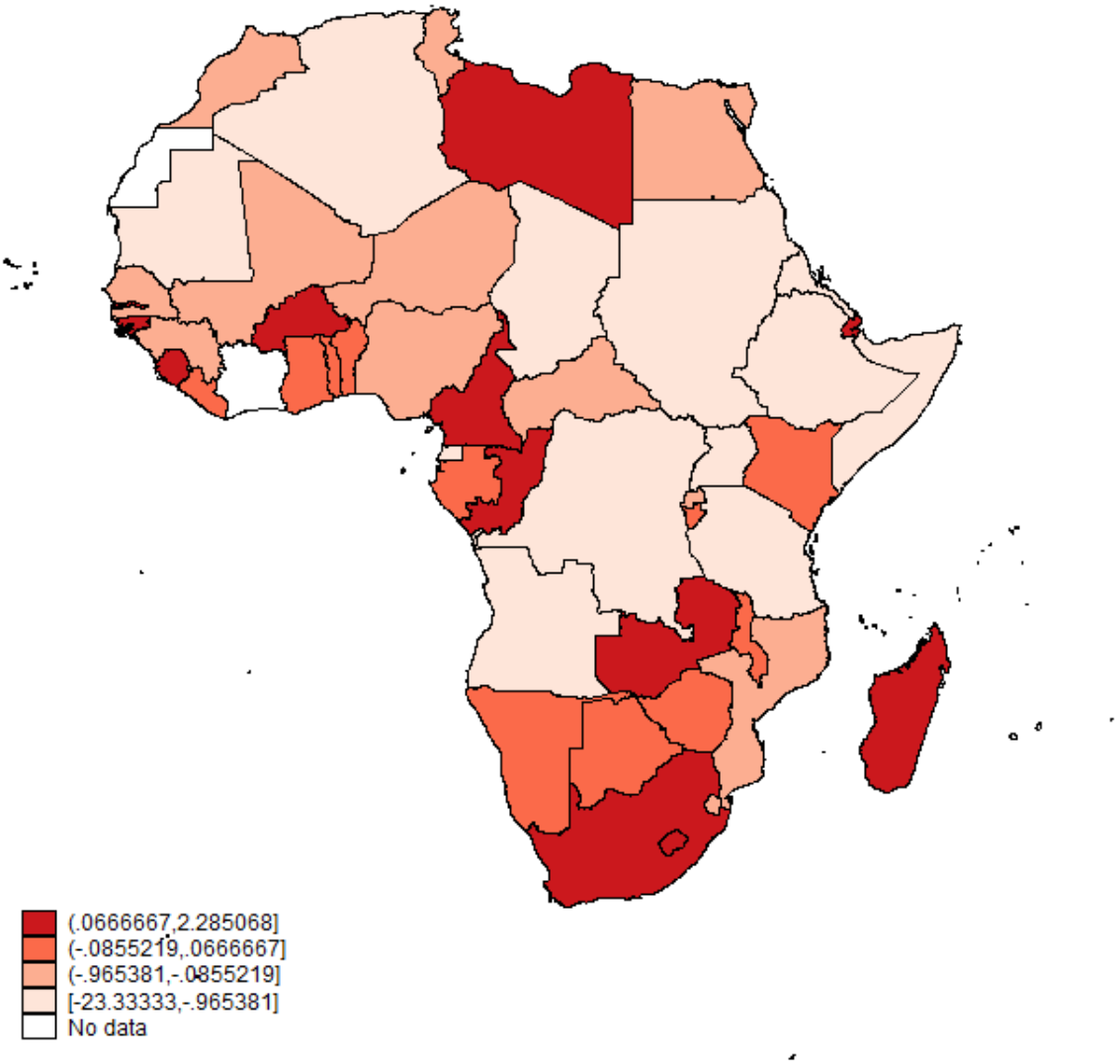


Figure 3.2: Severity of the terrorist attacks in Burkina Faso, Mali and Nigeria (in changes between 2010 and 2012)

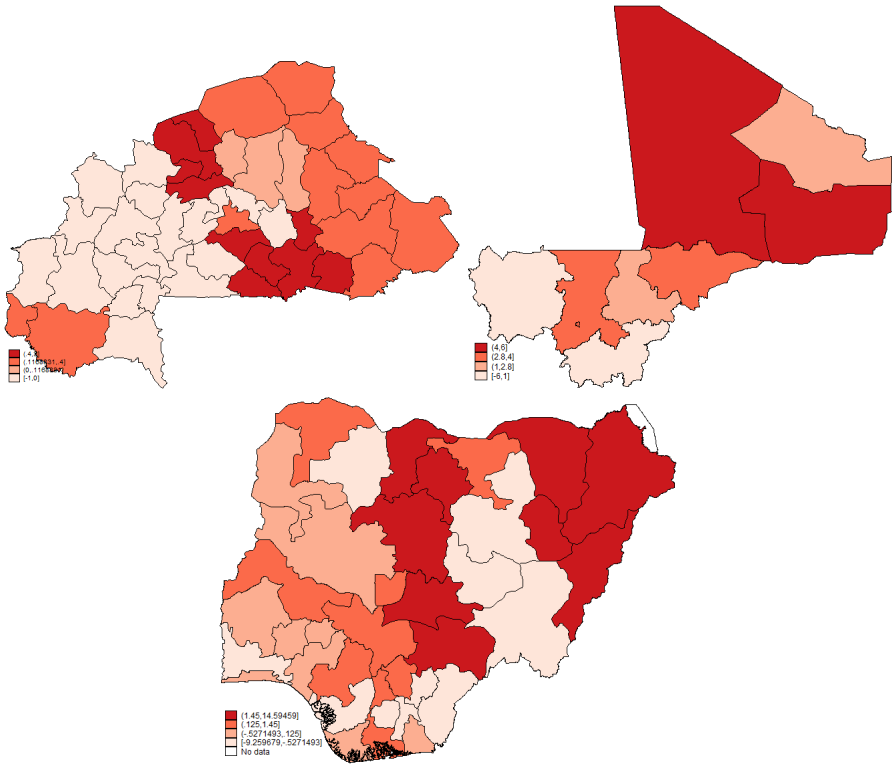


Figure 3.3: Terrorist Attacks

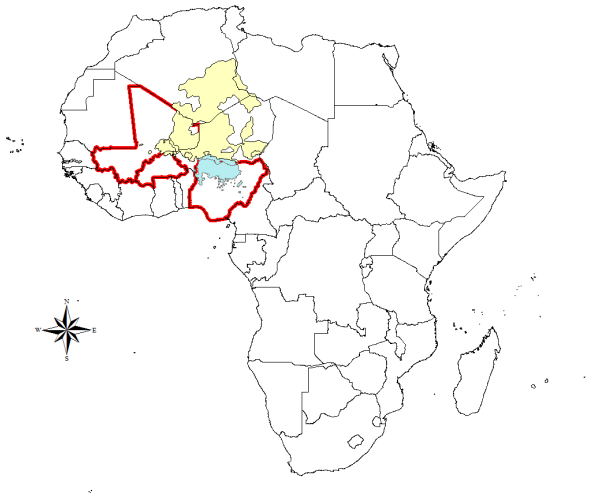
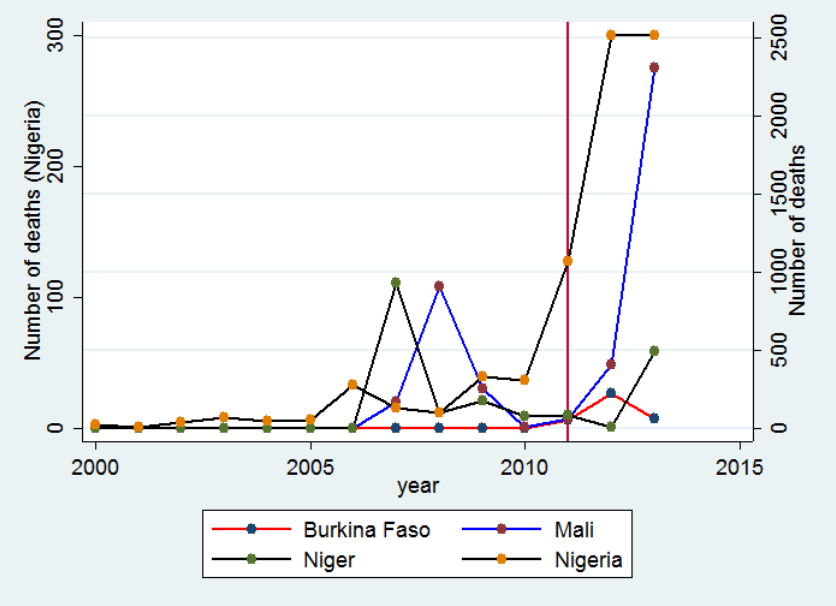


Figure 3.4: Terrorist Attacks



Nigerian society and to the concentration of the wealth in the predominantly Christian South of the country.

It is important to point out that our theory does not require that an insurgency is necessarily driven by the ethnic cleavages. The only requirement that we impose is that insurgency is perceived as a sign of state weakness (which is almost tautological because the existence of an active insurgency means that the state is not fully capable to maintain a monopoly on violence). Two types of insurgences that we study here presumably happened for a different sets of reasons: Tuareg insurgents were seeking independence of their ethnic homelands, while Boko Haram members were animated by the religious zeal and the perceived economic injustice. We have selected those insurgencies because they happened in the countries where we have reliable data on the salience of ethnic and national identities in the years *before* the insurgencies and thus can explore the potential effects of the insurgencies. It is also important that those insurgencies covered only several regions in their respective countries for this allows us to make within-country comparisons. Obviously, those regions were not in any way "random". We explore this assumption is the following sections.

4 Data

For this paper we use several datasets. The individual data before and after the start of insurgency of 2011-2012 come from the 4th and the 5th waves of Afrobarometer surveys[7]. As we can use only data for those countries, that are present in both waves – before and after the start of the insurgency – we restrict ourselves with study of Burkina Faso, Mali and Nigeria.

[7] `www.afrobarometer.org`

The main variable of interest is the national identity (NI_{it}). We construct it as an indicator variable by using the question about whether a respondent considers "national identity" their primary identity: if a respondent answers that they consider only "national identity" their primary identity, we assign the value of 1 to the NI_{it} and 0 otherwise.

The data for the number of terrorist attacks and fatalities due to the insurgencies are taken from the Global Terrorist Database (GTD [2015]) and Armed Conflict Location & Event Data Project[8].

Mali's Afrobarometer samples contains information about six South-Western regions (out of nine regions) of Mali as three northeastern regions are constantly dangerous because of the presence of *Al-Queda*. Tuareg activity was noted in four out of six regions[9], while two of them (*Bamako* and *Kayes*) can be considered as a control group. As for Burkina Faso, there are eight regions, and for the treatment group we use the northern region of *Sahel* where *Tuaregs* do live.

For the analysis, we use 2010 as a base year and 2012 as a treatment year. As can be seen on the figure 3.4 there was only one violent incident in 2009, and most of the attacks started in the late 2010 (Umar [2011]). Aggression was gradually rising until it peaked in the January 2012. On Figure 3.4 Nigerian data are depicted with the black lines and green dotes. We use 2010 as a base year, and 2012 as a treatment year: as can be seen there were a sharp increase in the number of killed and wounded in 2011, and even sharper increase in 2012. There are 36 states and one federal capital territory in Nigeria, and as a treatment group we will use all states where Boko Haram made at least one attack since 2010[10].

The historical data about the *Tuareg's* homelands are taken from the Geo-referencing of ethnic groups (GREG), that is a digitalized map of the Soviet ethnographic atlas "*Atlas Narodov Mira*".

5 Empirical Strategy and Identification

The main prediction from our model is that an insurgency has a negative causal effect on the national self-identification. To test this prediction we perform a series of statistical tests that impose different assumptions on the data-generating process. For the purpose of testing our theory, a random assignment of an insurgent activity to different geographical regions would provide the most convincing evidence. Such experimental assignment would be unethical. In this paper, we thus rely on the observational data (described above) and a series of estimation strategies. We start with a naive OLS estimation to find if we see the correlation between insurgency and national identity. Then we perform a difference-in-difference estimation where we assign historical ethnic homelands of *Tuaregs* and *Hausa/Fulani* to the hypothetical "treatment group" and other regions – to the hypothetical "control group'. This estimation allows us to rule out the existence of an unobserved constant region-specific factor that can bias our estimates. Finally, we perform an instrumental

[8]We use ACLED Version 6 from `http://www.acleddata.com/`.

[9]It is worth to note that we mark regions as treated if there were terrorist attacks conducted by *Tuaregs* in 2011-2013 (according to GTD [2015]). There were no *Tuaregs* attack in 2010.

[10]The data is taken from GTD [2015]. The list of those states is as follows: Adamawa, Kano, Gombe, Taraba, Bauchi, Plateau, Kaduna, Niger, Kogi, Abuja, Nasarawa, Zamfara, Sokoto, Katsina, Jigawa.

variable estimation, where we use an interaction of time and a status of ethnic homelands of a rebellious group to predict the instances of civil conflicts. Given the nature of our data, all methods we employ rely on certain untested assumptions, but because our results are robust across different methods, we conclude that, to the best of our knowledge, our theory is corroborated by the evidence.

We start with a ''naive" linear regression equation estimated via OLS:

$$NI_{irt} = \alpha + \delta_{\text{naïve}} Insurgency_{rt} + \mathbf{X'}_{irt}\gamma + \mu_r + \varepsilon_{irt}, \tag{5.1}$$

where as a dependent variable we use a dummy $NI_{irt} = 0$ if respondent i in region r has identified herself with her ethnic group at time $t \in \{2010, 2012\}$ and equal to 1 if she identifies herself with the nation. Variable $Insurgency_{rt}$ is a continuous measure of terrorist activity in region r at time t, $\mathbf{X'}_{irt}$ is the matrix of individual controls such as total years of education, age, religion, dummy for rural area, household's wealth index, religion and ethnicity, and μ_r is a region fixed effect.[11] The coefficient of interest is $\delta_{\text{naïve}}$, and we expect it to be negative. As the treatment is on the regional level, we cluster errors on the regional level as well.

Heroic assumptions are required for estimates of $\delta_{\text{naïve}}$ in equation 5.1 to be interpreted as the causal effect of civil conflict, as the instance of conflict could be correlated with omitted factors affecting both insurgency as well as national identity. It is important to note though, that as we have a region fixed effect in the regression, we already control for those factors that additively-linear and constant in time. To control for a time-varying factor, we add a year fixed effect and an interaction of the fixed effect of insurgency and a year fixed effect. This interaction term now becomes our main coefficient of interest:

$$NI_{irt} = \alpha + \beta_1 Insurgency_{rt} + \beta_2 POST_t + \tilde{\delta}_{\text{naïve}} \left(POST_t \times Insurgency_{rt}\right) + \mathbf{X'}_{irt}\gamma + \mu_r + \varepsilon_{irt}, \tag{5.2}$$

where as a dependent variable we use a dummy $NI_{irt} = 0$ if respondent i in region r has identified herself with her ethnic group at time $t \in \{2010, 2012\}$ and equal to unity if she identifies herself with the nation. Variable $POST_t$ is a dummy equal to unity if $t = 2012$, $Insurgency_r$ is a continuous measure of terrorist activity in region r at time t, X_i is the matrix of individual controls such as total years of education, age, religion, dummy for rural area, household's wealth index , religion and ethnicity, and μ_r is a region fixed effect. The coefficient of interest is $\tilde{\delta}_{\text{naïve}}$, and we expect it to be negative. As the treatment is on the regional level, we cluster errors on the regional level as well.

In the specification above, we are interested in $\widehat{\tilde{\delta}_{\text{naïve}}}$ that is the effect of insurgency when the civil conflict had happened.[12] Thus we use two types of variation. First, it is geographical variation: whether the particular region within a country experienced insurgency. And second, the temporal variation: whether the particular wave of the survey happened after the increase in violence.

[11] If in this or the following specifications we use the following assets instead of the wealth index: house phone, motor vehicle, television, radio, electricity, and type of access to water, results remain virtually unchanged.

[12] Results of regression with other proxies for insurgency will be provided in Robustness section.

One of the potential sources of bias is that the effect, decrease of national identity, might come not from the mechanism that we theorize about, the deteriorating belief about the strength of the state, but from a more trivial explanation: the rebellious group might withdraw the support for the nation. To make sure that this effect is not driving the results we exclude responses from the members of Tuareg group from our analysis, and we also exclude two regions directly controlled by *Boko Haram.*

An alternative strategy is to exploit the plausibly exogenous start of civil conflict described in Section 3 which increased the severity of civil conflict by making insurgency more likely. Assuming that the only outcome-relevant change before and after the *Tuareg* and *Boko Haram* rebellion, conditional on $\mathbf{X'}_{irt}$, was the increase in insurgency, we can use a difference-in-difference (DD) framework to specify the following regression:

$$NI_{irt} = \alpha + \beta POST_t + \delta_{DD} \left(POST_t \times Homeland_r \right) + \mathbf{X'}_{irt}\gamma + \mu_r + \varepsilon_{irt}, \tag{5.3}$$

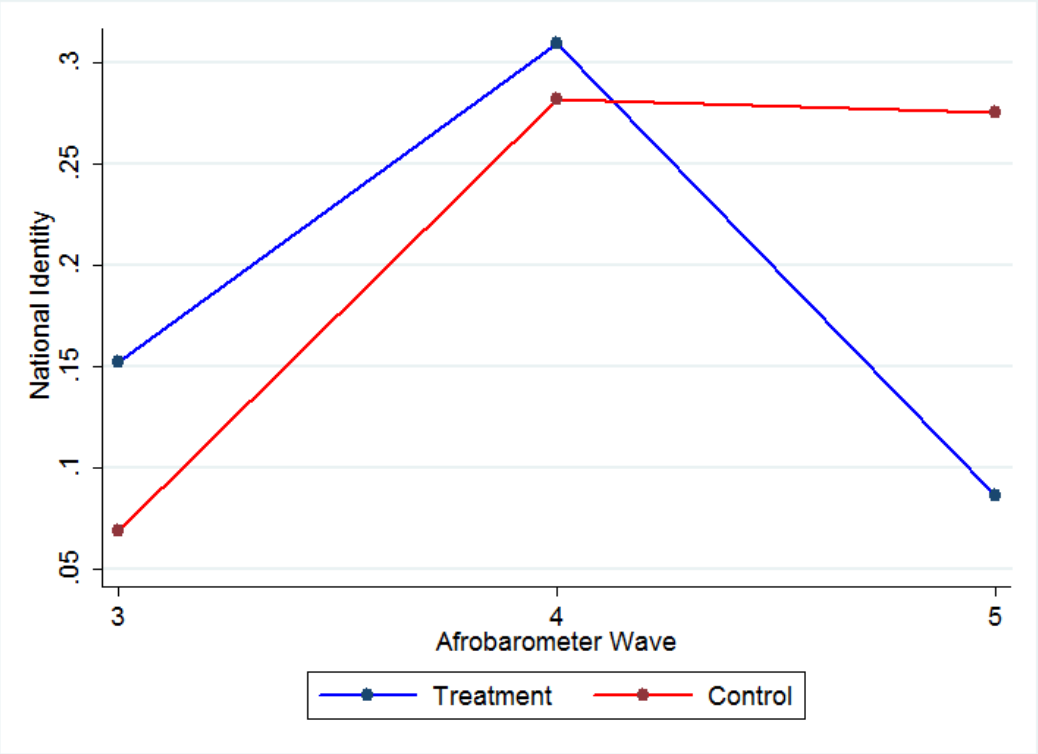
where as a dependent variable we use a dummy $NI_{irt} = 0$ if respondent i in region r has identified herself with her ethnic group at time $t \in \{2010, 2012\}$ and equal to unity if she identifies herself with the nation. Variable $POST_t$ is a dummy equal to unity if $t = 2012$, $Homeland_r$ is a dummy, if the region r is a homeland region for the ethnic group of *Tuareg*, or *Hausa/Fulani*, X_i is the matrix of individual controls such as total years of education, age, religion, dummy for rural area, household's wealth index , religion and ethnicity, and μ_r is a region fixed effect. Again, as the treatment is on the regional level, we cluster errors on the regional level as well. Here omit the $Homeland_r$ dummy as now it is collinear with region fixed effects.

It is important to note that the assignment regions to the "treatment" and "control" groups is not based on the levels of actual increase in violence, but on the *potential* exposure to the insurgent activity. In particular, in case of *Tuareg* rebellion, we assign Tuareg homelands to the treatment group, and in case The main assumption here is that, as those ethnic groups settled in those territories for hundreds of years long before the existence of any national borders, and the only way they can affect contemporary national identity choices is through higher probability of having a rebellious population nowadays.

One of the key assumptions of the difference-in-difference approach is that the potential outcomes for the control group and for the treatment group are the same. This assumption cannot be tested directly, however one of the ways to explore its plausibility is to look at trends in the dependent variable before the treatment within the treatment group and within the control group. If those trends are parallel, and the changes in trends coincide with the time of the treatment, it should increase our confidence, that the usual challenges to causal identification might be alleviated in this case.

Figure 5.1. shows the trends in the national identity. Blue lines represent the treatment group, and red line represents the control group. We see that before the insurgency (round 4 of the survey), the average levels of national identity have been increasing in all the regions, but right after the start of the insurgencies, we see important changes. The level of national identity in the control group remains stable (around 27 percent), but the level of national identity in the treatment group goes down rapidly (from 30 percent to 9

Figure 5.1: Pre-treatment trends in national identity



percent).

Under the assumption that national identification used to exhibit similar trends in all regions, we can think of the parameter δ_{DD} as measuring the effect of civil conflict when in a regions affected by rebellion. However, if for example the insurgency increase over time for reasons that could themselves affect the outcome, the underlying identification assumption is easily violated. For example, a shift towards Islamic fundamentalism in treated regions over time could induce interpretations of the DD estimate other than one related to the effect of civil conflict. For this reason we further exploit historical homelands of ethnic groups involved in the civil conflict in an instrumental variable (IV) framework. The IV's first-stage regression is specified as follows:

$$Insurgency_r = \pi_0 + \pi_1 POST_t + \pi_2 \left(Homeland_r \times POST_t\right) + \mathbf{X'}_{irt}\tilde{\gamma} + \mu_r + \nu_{irt},$$

And the second stage looks like:

$$NI_{irt} = \alpha + \delta_{\text{naïve}} \widehat{Insurgency}_{rt} + \mathbf{X'}_{irt}\gamma + \mu_{rt} + \varepsilon_{irt}, \tag{5.4}$$

where as a dependent variable we use a dummy $NI_{irt} = 0$ if respondent i in region r has identified herself with her ethnic group at time $t \in \{2010, 2012\}$ and equal to unity if she identifies herself with the nation.

Variable $POST_t$ is a dummy equal to unity if $t = 2012$, $Homeland_r$ is a dummy, if the region r is a homeland region for the ethnic group of *Tuareg*, or *Hausa/Fulani*; $Insurgency_r$ is a continuous measure of terrorist activity in region r at time t; X_i is the matrix of individual controls such as total years of education, age, religion, dummy for rural area, household's wealth index , religion and ethnicity, and μ_r is a region fixed effect. Errors are clustered on the regional level as well.

6 Results

Results of the OLS estimations are presented in Table 1. In Columns I we present results of the naive OLS specification for with the full set of demographic controls and region fixed effects. While result is negative, it is insignificant. The reason is, terrorist attack per se may be correlated with the national identity if national identity and insurgency show the similar trends. To take this into account, we include dummy for the year after the start of civil conflict and interaction of the severity of insurgency with the time dummy. Results are presented in Column II: the interaction term is negative and significant, suggesting, that insurgency during the civil conflict is negatively related to the choice of national identity. In particular, one standard deviation change in the severity of insurgency decreases probability of a person to associate herself with a nation by 9.5%.

In Column III we add ethnicity fixed effects, however the coefficient of interest does not change significantly.

To show that our results are not driven by sampling of the Afrobarometer survey, in Columns IV and V we present results of similar regression specifications but on the aggregated by region sample. In particular, Column IV represents the same specification as Column I: similarly, the coefficient is negative, while significant. We add dummies for countries, instead of region dummies, and use robust standard errors.[13] Column V repeats the specification in Column II, with all coefficients having the same signs and being significant. The result suggest that one standard deviation in insurgency after the civil war started decreases share of people who associated themselves with the nation by 35.1% of the standard deviation.

As we would expect, we do not capture the causal effect due to endogeneity issues. First, measurement error of our main explanatory variable, that appear due to under-reported incidents of the civil conflict may attenuate our results. Second, more problematic bias can be caused by unobserved heterogeneity caused by the fact, that *Tuareg* and *Hausa\Fulani* ethnic groups currently dwell in those regions where insurgency and perception of national identity correlate with some institutional or cultural unobservables. Third, reverse causality concerns remain, as adverse shock of national identity due to the civil conflict can increase the insurgency.

[13] Clustering by countries (three clusters) yields standard errors that are smaller than robust one, and we prefer to show them as the most conservative. Clustered by country standard errors are available upon request.

Table 1: Civil conflict and national identity: OLS

	I	II	III	IV	V
VARIABLES	Dependent variable: National identity				
Insurgency	-0.0133	0.0270	0.0279	-0.00914**	0.01601
	(0.0111)	(0.0173)	(0.0167)	(0.00382)	(0.01221)
After civil war		-0.0407	-0.0675		-0.0592**
		(0.0417)	(0.0447)		(0.0257)
Terrorist attacks ×		-0.0340**	-0.0326**		-0.0256**
After civil war		(0.0152)	(0.0143)		(0.0125)
Socioeconomic controls	✓	✓	✓	✓	✓
Region FE	✓	✓	✓	×	×
Country FE	✓	✓	✓	✓	✓
Ethnicity FE	×	×	✓	×	×
Sample		Individual		Aggregated	
Observations	9,349	9,556	9,349	225	225
R-squared	0.169	0.156	0.175	0.411	0.451

Notes: All columns include constant. The following variables are included as controls in columns I-V: age, age squared, gender dummy, religion dummy (christians, muslim or traditional), dummy for positive change in living conditions, dummy for unemployment, dummy for dayly access to news, wealth index. Controls for columns IV and V are aggregated. In columns I-III robust clustered by region standard errors in parentheses. 60 clusters. In columns IV and V robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

In the Table 2 we present our difference in difference estimates based on the ethnic homelands of the rebellious tribes (historical ethnic homelands of *Tuaregs* (in case Burkina Faso and Mali) and *Hausa\Fulani* ethnic groups (in case of Nigeria)) rather than actual insurgency (specification 5.3). In Column I we present results of the regression without covariates. The difference-in-difference estimator tells us, that civil war in those territories decreases probability of accepting national identity by 17.6% in historical territories of affected tribes. By adding covariates in Column II we have the same result, such as the coefficient $\widehat{\delta_{DD}}$ does not change, that is a good sign, as by our assumption of the difference-in-difference estimation, the treatment should be uncorrelated to the covariates. Further, in Column III we add ethnicity fixed effects, that should explain a lot in the individual unobserved heterogeneity. The coefficient of interest decreases in magnitude, such as civil war in homelands of *Tuaregs* and *Hausa\Fulani* decreases probability of accepting national identity by 15.1%.

Finally, we estimate the same model specification with the data aggregated on the regional level in Columns IV and V. Both results in negative and significant coefficient of the interaction term: civil war in homelands of *Tuaregs* and *Hausa\Fulani* decreases share of people accepted national identity by 48.2%

of its standard deviation. Figure6.1 shows the residual plot of the national identity and treatment status interacted with time dummy.

Table 2: Civil conflict and national identity: Difference-in-difference estimation

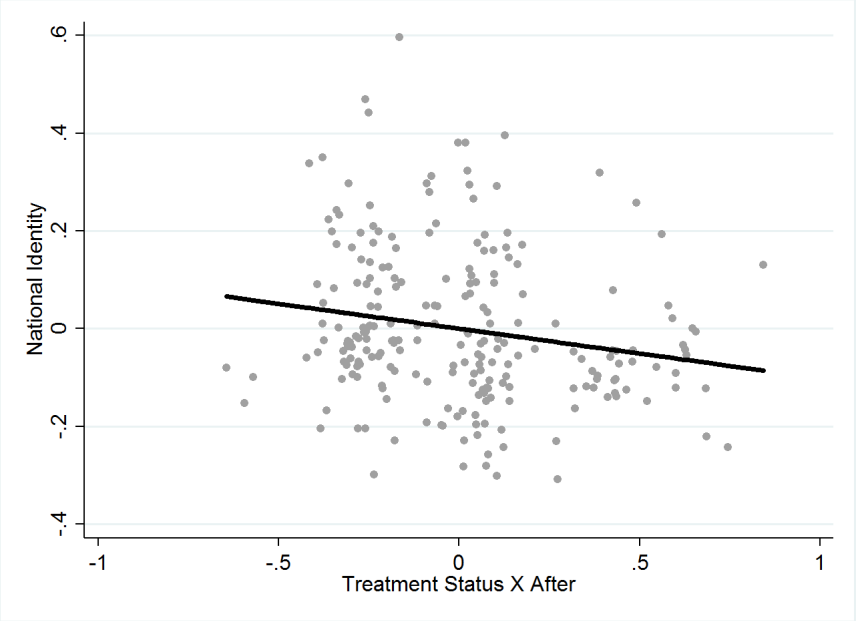
	I	II	III	IV	V
VARIABLES	Dependent variable: National identity				
After civil war	-0.0186	-0.0214	-0.0434	-0.0754***	-0.0547**
	(0.0349)	(0.0340)	(0.0387)	(0.0228)	(0.0250)
Homelands ×	-0.176**	-0.172**	-0.151*	-0.0610**	-0.102***
After civil war	(0.0770)	(0.0761)	(0.0793)	(0.0256)	(0.0328)
Socioeconomic controls	×	✓	✓	×	✓
Region FE	✓	✓	✓	×	×
Country FE	✓	✓	✓	✓	✓
Ethnicity FE	×	×	✓	×	×
Sample	Individual			Aggregated	
Observations	9,556	9,556	9,349	225	225
R-squared	0.158	0.160	0.178	0.399	0.453

Notes: All columns include constant. The following variables are included as controls in columns I-V: age, age squared, gender dummy, religion dummy (christians, muslim or traditional), dummy for positive change in living conditions, dummy for unemployment, dummy for dayly access to news, wealth index. Controls for columns IV and V are aggregated. In columns I-III robust clustered by region standard errors in parentheses. 60 clusters. In columns IV and V robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

One can also interpret these results as a reduced for the specification 5.4 as we assume that the historic homelands interacted with time fixed effects can affect changes in national self-identification only through the actual increase in violence. To explore this idea more thoroughly, we employ our IV framework, following specification 5.4. First, in Columns I we present reduced form individual-level regression. Both instruments, are significant, in predicting the insurgency, such as more severe insurgency happens after the start of the civil conflict and especially in the ethnic homelands of the affected tribes. The F statistics of excluded instruments is equal to 13.8 that is marginally bigger than 11.59 - the proposed critical value for the weak instrument defined by Stock and Yogo [2005] for one endogenous regressor and two instruments. At the same time, Anderson-Rubin test suggest, that the instrument is not weak on 1% confidence level. While the instruments are not weak, the partial R squared is quite high (0.35), suggesting that two instruments solely explain third of the variation in the endogenous variable. This can potentially make instruments capture some of the endogeneity of the instrumented variable. Discussion of this issue and sensitivity checks that will alleviate this concern will be presented later in the Robustness section.

Results of the second stage presented in Column II. As we see, the coefficient of interest is negative and

Figure 6.1: Difference-in-difference estimation, residual plot: National identity and the interaction term (aggregated)



significant, and slightly bigger in magnitude than OLS specifications in Table 1: one standard deviation change in the severity of insurgency decreases probability of a person to associate herself with a nation by 11.3%. This suggests that reverse causality and unobserved heterogeneity concerns were less of an issue.

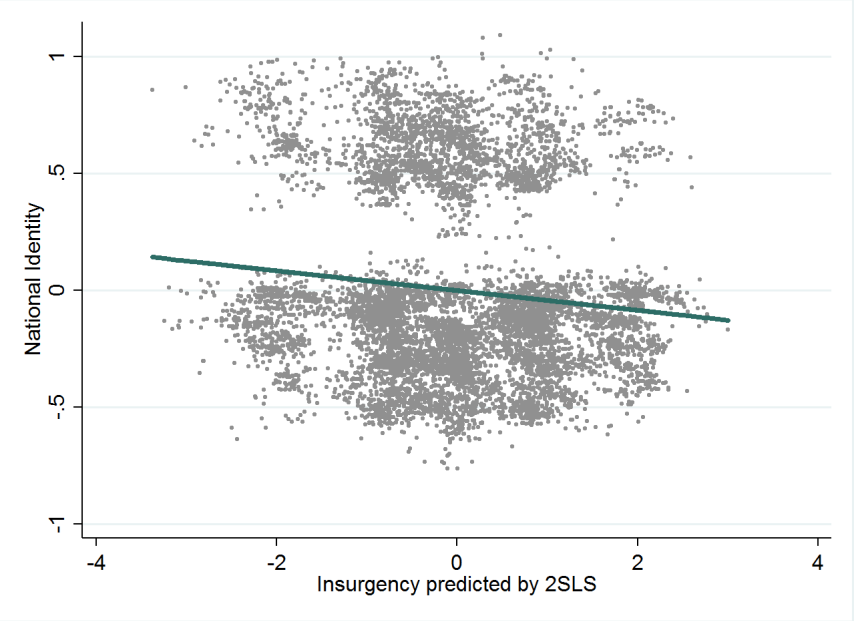
In Columns III and IV we present the results of the similar specification on the aggregated data. The first stage is reported in the Column III. Aggregated data specification yields stronger first-stage F-statistics (35.1), and slightly lower partial $R^2 = 0.31$. Similarly, Anderson-Rubin p-value is below 1% confidence level. IV estimate in Column IV is 24% bigger than the OLS one, suggesting that one standard deviation in insurgency after the civil war started decreases share of people who associated themselves with the nation by 43.6% of the standard deviation. Figure6.2 shows the residual plot from the individual-level estimates.

Table 3: Civil conflict and national identity: IV estimation

	I	II	III	IV
VARIABLES	Second stage dependent variable: National identity			
	First stage	Second stage	First stage	Second stage
After civil war	1.690**		0.468	
	(0.750)		(0.404)	
Homelands ×	2.283*		4.160***	
After civil war	(1.148)		(0.633)	
Insurgency		-0.0422***		-0.0318***
		(0.0160)		(0.00746)
Socioeconomic controls	✓	✓	✓	✓
Region FE	✓	✓	×	×
Country FE	✓	✓	✓	✓
Ethnicity FE	✓	✓	×	×
Partial R-squared	0.351		0.313	
F-stat of excluded instruments	13.774		35.123	
Prob ¿ F	0.000		0.000	
Anderson-Rubin p-value	0.006		0.000	
Sample	Individual		Aggregated	
Observations	9,349	9,349	225	225
R-squared	0.710	0.15	0.423	0.33

Notes: Columns I and III are estimated with OLS. Columns II and IV are estimated with 2SLS. Instrumented variable in columns II and IV is insurgency. Instruments in columns II and IV are dummy for year after civil conflict and ethnic homelands dummy. All columns include constant. The following variables are included as controls in columns I-IV: age, age squared, gender dummy, religion dummy (christians, muslim or traditional), dummy for positive change in living conditions, dummy for unemployment, dummy for dayly access to news, wealth index. Controls for columns III and IV are aggregated. In columns I-II robust clustered by region standard errors in parentheses. 60 clusters. In columns III and IV robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 6.2: IV estimation, residual plot: National identity and the interaction term (individual-level)



However, the results above may not hold if assumptions of instrumental variable estimation do not hold. In the Table 4 we check if the effect is indeed comes through compliers: peoples who stop associate themselves as a nation in regions - ethnic homelands of the ethnic groups involved in the civil conflict. As can be seen, both individual level specification (Columns I and II) and aggregated level specification (Columns III and IV) yield weak first stage and insignificant coefficient of interest on the second stage. This fact indicates that the effect is indeed coming through the compliers and not through other means.

Another important concern regarding results of the IV estimate is violation of the exclusion restrictions. While it is hardly possible to prove absence of its violation, it is hard to believe that ethnic homelands that were established hundreds of years ago affected something that had happened exactly between 2010 and 2012. Nevertheless, we take this concern seriously, and in Section 7 will employ method developed in Conley et al. [2012] and relax exogeneity assumption of the instruments to show that results are still hold.

Table 4: Civil conflict and national identity: IV estimation (no compliers)

	I	II	III	IV
VARIABLES	Second stage dependent variable: National identity			
	First stage	Second stage	First stage	Second stage
After civil war	1.686**		1.109***	
	(0.751)		(0.375)	
Homelands ×	-1.896**		-0.577**	
After civil war	(0.871)		(0.246)	
Insurgency		-0.0290		-0.0256
		(0.0293)		(0.0245)
Socioeconomic controls	✓	✓	✓	✓
Region FE	✓	✓	×	×
Country FE	✓	✓	✓	✓
Ethnicity FE	✓	✓	×	×
Partial R-squared	0.201		0.102	
F-stat of excluded instruments	2.522		4.471	
Prob ¿ F	0.091		0.013	
Anderson-Rubin p-value	0.004		0.487	
Sample	Individual		Aggregated	
Observations	7,755	7,755	169	169
R-squared	0.631	0.16	0.220	0.42

Notes: Columns I and III are estimated with OLS. Columns II and IV are estimated with 2SLS. Instrumented variable in columns II and IV is insurgency. Instruments in columns II and IV are dummy for year after civil conflict and ethnic homelands dummy. All columns include constant. The following variables are included as controls in columns I-IV: age, age squared, gender dummy, religion dummy (christians, muslim or traditional), dummy for positive change in living conditions, dummy for unemployment, dummy for dayly access to news, wealth index. Controls for columns III and IV are aggregated. In columns I-II robust clustered by region standard errors in parentheses. 60 clusters. In columns III and IV robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Results of this section show that civil conflict in Mali, Burkina Faso and Nigeria had negative causal effect on peoples national identity.

7 Robustness and Sensitivity Checks

In this section we present a set of robustness checks, that intend to show that our results are not driven by data artifacts and can indeed be causally interpreted. First, we will show that results are robust to various proxies of insurgency. Second we will show that the results are not driven by one of the countries in

the sample or some particular regions. Third, we will report placebo test, that show that ethnic homelands does not affect other important variables. Finally, following Conley et al. [2012] we will relax exogeneity assumptions of our instruments and show that the result still holds.

Table 5: Civil conflict and national identity: IV estimation

	I	II	III	IV	V
VARIABLES	Second stage dependent variable: National identity				
Insurgency	-0.0290**	-0.0542**	-0.0983***	-0.172***	-0.199***
	(0.0114)	(0.0237)	(0.0355)	(0.0662)	(0.0735)
Controls	✓	✓	✓	✓	✓
Region FE	✓	✓	✓	✓	✓
Ethnicity FE	✓	✓	✓	✓	✓
Partial R-squared	0.422	0.278	0.379	0.462	0.496
F-stat of excl inst.	16.380	6.618	14.694	22.170	29.403
Prob ¿ F	0.000	0.003	0.000	0.000	0.000
Sample	Individual-level				
Data source	GTD	ACLED	Both	ACLED	GTD
Insurgency proxy	$\frac{\sum fatalities+\sum wounded}{\sum \text{number of incidents}}_{rt}$	$\frac{\sum fatalities}{\sum \text{number of incidents}}_{rt}$	PCA of both $\frac{\sum fatalities}{\sum \text{number of incidents}}_{rt}$	Dummy for insurgency	
Observations	9,349	9,349	9,349	9,349	9,349
R-squared	0.16	0.15	0.16	0.17	0.16

Notes: Columns I and III are estimated with OLS. Columns II and IV are estimated with 2SLS. Instrumented variable in columns II and IV is insurgency. Instruments in columns II and IV are dummy for year after civil conflict and ethnic homelands dummy. All columns include constant. The following variables are included as controls in columns I-IV: age, age squared, gender dummy, religion dummy (christians, muslim or traditional), dummy for positive change in living conditions, dummy for unemployment, dummy for dayly access to news, wealth index. Controls for columns III and IV are aggregated. In columns I-II robust clustered by region standard errors in parentheses. 60 clusters. In columns III and IV robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

In Tables 5 and 6 we report results for the specifications similar to those in Table 3 but with different proxies for insurgency. First, in Column I I use the measure that takes into account not only fatalities but wounded individuals for computing the severity of insurgency. In Column II I use the same proxy as in Column II of Table 3 but computed with the ACLED data instead of GTD. In Column III I use the first principal component of the GTD and ACLED measures used in Column II of Tables 3 and 5. In Columns IV and V I employ binary measure of insurgency, computed as the dummy if the total number of fatalities in the region r in year t is above the median.[14] As can be seen, the average F statistics of the excluded instruments

[14]The median for the GTD is 0 fatalities, while for ACLED, it is 8. Using other thresholds do not significantly change the

is 17,9 and varies from 6.6 to 29.4 in individual-level specifications. The specification in Column II yields F statistics below one that Stock and Yogo [2005] suggests, however, the first stage is still significant on 1% level, and Anderson-Rubin p-value is below 0.01. All second stage coefficients has negative and significant coefficient supporting our findings. Columns V-X contain specifications similar to those in Columns I-V but for the aggregated data. Again, F statistics of the first stage is pretty strong, ranging from 18.8 to 84.8, and the coefficient of interest is negative and significant.

Table 6: Civil conflict and national identity: IV estimation, continuation

VARIABLES	VI	VII	VIII	IX	X
Insurgency	-0.0190***	-0.0342***	-0.0634***	-0.239***	-0.190***
	(0.00478)	(0.00973)	(0.0160)	(0.0621)	(0.0413)
Controls	✓	✓	✓	✓	✓
Region FE	✓	✓	✓	✓	✓
Ethnicity FE	✓	✓	✓	✓	✓
Partial R-squared	0.270	0.207	0.219	0.404	0.305
F-stat of excl inst.	34.757	18.879	30.534	84.821	28.804
Prob ¿ F	0.000	0.000	0.000	0.000	0.000
Sample			Aggregated		
Data source	GTD	ACLED	Both	ACLED	GTD
Insurgency proxy	$\frac{\sum fatalities+\sum wounded}{\sum \text{number of incidents}}_{rt}$	$\frac{\sum fatalities}{\sum \text{number of incidents}}_{rt}$	PCA of both $\frac{\sum fatalities}{\sum \text{number of incidents}}_{rt}$	Dummy for insurgency	
Observations	225	225	225	225	225
R-squared	0.28	0.21	0.32	0.38	0.30

Notes: Columns I and III are estimated with OLS. Columns II and IV are estimated with 2SLS. Instrumented variable in columns II and IV is insurgency. Instruments in columns II and IV are dummy for year after civil conflict and ethnic homelands dummy. All columns include constant. The following variables are included as controls in columns I-IV: age, age squared, gender dummy, religion dummy (christians, muslim or traditional), dummy for positive change in living conditions, dummy for unemployment, dummy for dayly access to news, wealth index. Controls for columns III and IV are aggregated. In columns I-II robust clustered by region standard errors in parentheses. 60 clusters. In columns III and IV robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Another possible concern is that our results may be driven by one particular civil conflict, as in our estimates we use two: Tuareg rebellion and Boko Haram insurgency. To take this into account we split the sample into two: Mali and Burkina Faso, and Nigeria. By doing this we have too few observations

results.

to estimate the aggregated specifications: while having correct signs, those results become insignificant. However, in Table 7 we report results of the individual level specifications.

First, In Columns I and II we present difference-in-difference estimates for Mali and Burkina Faso samples for no controls and full set of the controls specifications. The coefficients of interest is negative, significant, and almost twice larger than the coefficients of the corresponding difference-in-difference specifications in Table 2. And indeed if we look at the Nigerian sample in Columns V and VI the magnitude of the coefficients is twice as small, and significant (on 1% level) only for the full control specification), suggesting, that ethnic homelands are less important as a determinant of insurgency in Nigeria than in Mali and Burkina Faso.

Columns III and IV contain results of the 2SLS estimation for the sample without Nigeria. The resulting second stage coefficient is again twice as large as the one in the full sample. Although, the first stage is marginally strong, with F-statistics equal to 4.9 and Anderson-Rubin p-value of 0.056. Results for the Nigerian sample are similar: F-statistics equal to 4.9 and Anderson-Rubin statistics significant on 1%, and significant, but smaller in magnitude second stage coefficient. These results suggest, that while both civil conflicts had affect on the national identity, the effect we estimate for the *Tuareg* rebellion is much stronger than the effect of *Boko Haram*. Howeve, this can be explained by the fact, that *Boko Haram* movement was not exclusively supported by *Hausa/Fulani* ethnic group, but other Muslim ethnic groups, thus making our identification weaker.

In addition, we test if our results are robust to alternative construction of the main dependent variable. In this case, the national identity (NI_{it}) dummy is equal to one if the respondent considers only "national identity" their primary identity or "more national identity" than "ethnic identity", and 0 otherwise. The results, of difference-in-difference and IV estimations are presented in Table 8. As can be seen all coefficients remain negative and significant.

Finally, in vein of Conley et al. [2012] we relax the exogeneity assumptions of the instruments and examine the bounds we are able to place on the true effect of insurgency on the national identity. The idea behind the method is simple: if in addition to exogenous and endogenous variables we add instruments (dummy for the year after the civil conflict had started and its interaction with ethnic homelands dummy) its coefficients (γ_1 and γ_2) required to be equal to zero according to standard IV estimation. However, by relaxing the constraint we can find the bounds for the IV estimate of insurgency (β). If one expects instruments to have direct or indirect positive effect on the choice of identity($\gamma_1 > 0$ and $\gamma_2 > 0$) we will underestimate the true effect of the insurgency on the national identity. This gives the maximum prior for γ. More challenging is to determine the minimum prior of γs. Thus we assume, that the maximum direct effect of instrument will be not bigger than the size of the biggest effect of one of the control covariates. The covariate with the biggest significant covariate (standardized) is the dummy if individual has access to the news on the daily basis. Thus we assign the minimum prior for both $\gamma_1 = \gamma_2 = -0.028$. Applying Conley et al. [2012], we find that the bounds on the strength of β are still below zero (at 95% confidence level):$[-0.049; -0.021]$ if we use our baseline measure of insurgency or $[-0.208; -0.068]$ if we use dummy for insurgency. Therefore, even allowing for imperfect exogeneity, the negative effect of insurgency on the national identity is confirmed.

Table 7: Civil conflict and national identity: IV estimation (without Nigeria)

VARIABLES	I	II	III	IV	V	VI	VII	VIII
	Dependent variable:							
	National identity		Insurgency	National identity	National identity		Insurgency	National identity
	Difference-in-difference		First stage	Second stage	Difference-in-difference		First stage	Second stage
After civil war	0.0580	0.0486	0.393		-0.173***	-0.172***	-0.257	
	(0.0464)	(0.0620)	(0.775)		(0.0324)	(0.0162)	(0.419)	
Homelands ×	-0.393**	-0.377**	2.567*		-0.0657	-0.0778***	4.063***	
After civil war	(0.186)	(0.162)	(1.280)		(0.0790)	(0.0254)	(1.307)	
Insurgency				-0.0928***				-0.0373*
				(0.0311)				(0.0198)
Socioeconomic controls	×	✓	✓	✓	×	✓	✓	✓
Region FE	✓	✓	✓	✓	✓	✓	✓	✓
Ethnicity FE	×	✓	✓	✓	×	✓	✓	✓
Partial R-squared			0.175				0.326	
F-stat of excluded instruments			4.907				4.919	
Prob ¿ F			0.018				0.013	
Anderson-Rubin p-value			0.0564				0.000	
Sample	Mali and Burkina Faso				Nigeria			
Observations	4,832	4,777	4,777	4,777	4,724	4,572	4,572	4,572
R-squared	0.064	0.086	0.519	0.07	0.149	0.160	0.740	0.09

Notes: Columns I and III are estimated with OLS. Columns II and IV are estimated with 2SLS. Instrumented variable in columns II and IV is insurgency. Instruments in columns II and IV are dummy for year after civil conflict and ethnic homelands dummy. All columns include constant. The following variables are included as controls in columns I-IV: age, age squared, gender dummy, religion dummy (christians, muslim or traditional), dummy for positive change in living conditions, dummy for unemployment, dummy for dayly access to news, wealth index. Controls for columns III and IV are aggregated. In columns I-II robust clustered by region standard errors in parentheses. 60 clusters. In columns III and IV robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Civil conflict and national identity: Alternative national identity measure

	I	II	III	IV	V	VI	VII
VARIABLES	Dependent variable: National identity						
	Difference-in-difference					Second stage	
After civil war	-0.0216	-0.0248	-0.0577	-0.0816**	-0.0954***		
	(0.0389)	(0.0377)	(0.0424)	(0.0327)	(0.0306)		
Homelands ×	-0.211**	-0.206**	-0.183**	-0.109***	-0.0787*		
After civil war	(0.0793)	(0.0790)	(0.0816)	(0.0406)	(0.0441)		
Insurgency						-0.0529***	-0.0334***
						(0.0178)	(0.00983)
Socioeconomic controls	×	✓	✓	×	✓	✓	✓
Region FE	×	✓	✓	×	✓	✓	✓
Ethnicity FE	×	✓	✓	×	×	✓	×
Partial R-squared						0.351	0.313
F-stat of excluded instruments						13.774	35.123
Prob ¿ F						0.000	0.000
Anderson-Rubin p-value						0.009	0.001
Sample		Individual		Aggregated		Individual	Aggregated
Observations	9,556	9,556	9,349	225	225	9,349	225
R-squared	0.129	0.133	0.149	0.097	0.359	0.12	0.22

Notes: Columns I and III are estimated with OLS. Columns II and IV are estimated with 2SLS. Instrumented variable in columns II and IV is insurgency. Instruments in columns II and IV are dummy for year after civil conflict and ethnic homelands dummy. All columns include constant. The following variables are included as controls in columns I-IV: age, age squared, gender dummy, religion dummy (christians, muslim or traditional), dummy for positive change in living conditions, dummy for unemployment, dummy for dayly access to news, wealth index. Controls for columns III and IV are aggregated. In columns I-II robust clustered by region standard errors in parentheses. 60 clusters. In columns III and IV robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

All robustness and sensitivity checks above suggest, that the our theory is indeed supported by the empirical data, and the civil conflict is damaging the creation of the national identity.

8 Conclusion

This paper looks at the connection between the national identity and insurgencies. In our model, we look at the formation of national identity as a problem of coordination anchored in the strength of the state. Because the insurgencies are the manifestations of state weakness, our model predicts that that the coordination on the national identity becomes harder if the citizens are exposed to an active insurgency.

We test the predictions of our model using the survey data from the the countries exposed to the recent insurgencies in Sub-Saharan Africa: *Tuareg* rebellion and *Boko Haram*. We find that in the regions with active insurgent campaigns, the national identity significantly decreased, while it remained stable in the regions not exposed to the rebellion.

Our study suggests that one of the many reasons why civil conflict is costly for growth is because it might erode national identity.

References

Daron Acemoglu, Garcia-JimenoCamilo, and James A Robinson. State capacity and economic development: A network approach. Technical report, National Bureau of Economic Research, 2014. 1

Paloma Aguilar Fernández and Carsten Humlebaek. Collective memory and national identity in the spanish democracy: The legacies of francoism and the civil war. *History & Memory*, 14(1):121–164, 2002. 1

Benedict Anderson. *Imagined communities: Reflections on the origin and spread of nationalism*. Verso Books, 2006. 1

Michal Bauer, Christopher Blattman, Julie Chytilová, Joseph Henrich, Edward Miguel, and Tamar Mitts. Can war foster cooperation? Technical report, National Bureau of Economic Research, 2016. 1

Timothy Besley and Torsten Persson. State capacity, conflict, and development. *Econometrica*, 78(1):1–34, 2010. 1

Hans Carlsson and Eric Van Damme. Global games and equilibrium selection. *Econometrica: Journal of the Econometric Society*, pages 989–1018, 1993. 2

Jean-Paul Carvalho. Coordination and culture. Technical report, University of Oxford, Department of Economics, 2010. 2

Timothy G Conley, Christian B Hansen, and Peter E Rossi. Plausibly exogenous. *Review of Economics and Statistics*, 94(1):260–272, 2012. 6, 7, 7

Melissa Dell and Pablo Querubin. Nation building through foreign intervention: Evidence from discontinuities in military strategies. Technical report, National Bureau of Economic Research, 2016. 1

Mark Dincecco and Mauricio Prado. Warfare, fiscal capacity, and performance. *Journal of Economic Growth*, 17(3):171–203, 2012. 1

Benn Eifert, Edward Miguel, and Daniel N Posner. Political competition and ethnic identification in africa. *American Journal of Political Science*, 54(2):494–510, 2010. 1

Peter P Ekeh. Colonialism and the two publics in africa: a theoretical statement. *Comparative studies in society and history*, 17(01):91–112, 1975. 1

Barbara Geddes. *Politician's dilemma: building state capacity in Latin America*. Number 25. Univ of California Press, 1994. 1

Ernest Gellner and John Breuilly. *Nations and nationalism*. Cornell University Press, 2008. 1

GTD. *National Consortium for the Study of Terrorism and Responses to Terrorism (START)*. Global Terrorism Database [Data file]. Retrieved from http://www.start.umd.edu/gtd, 2015. 5, 4, 9, 10

Eric J Hobsbawm. *Nations and nationalism since 1780: Programme, myth, reality*. Cambridge University Press, 2012. 1

Samuel P Huntington. *Who are we?: The challenges to America's national identity*. Simon and Schuster, 2004. 1

David Laitin. Identity in formation. *The Russian-speaking populations in the near abroad*, 1998. 1

Michael Mann. *The sources of social power: Volume 3, global empires and revolution, 1890-1945*, volume 3. Cambridge University Press, 2012. 1

Edward Miguel. Tribe or nation? nation building and public goods in kenya versus tanzania. *World Politics*, 56(03):328–362, 2004. 1

Stephen Morris and Hyun Song Shin. Unique equilibrium in a model of self-fulfilling currency attacks. *American Economic Review*, pages 587–597, 1998. 1, 2, 2, A.1

Stephen Morris and Hyun Song Shin. Coordination risk and the price of debt. *European Economic Review*, 48(1):133–153, 2004. 1, 2, 2, A.1

Mark E Neely. *Lincoln and the Triumph of the Nation: Constitutional Conflict in the American Civil War*. Univ of North Carolina Press, 2011. 1

Daniel N Posner. The political salience of cultural difference: Why chewas and tumbukas are allies in zambia and adversaries in malawi. *American Political Science Review*, 98(04):529–545, 2004. 1

Daniel N Posner. *Institutions and ethnic politics in Africa*. Cambridge University Press, 2005. 2

Amanda Lea Robinson. National versus ethnic identification in africa: Modernization, colonial legacy, and the origins of territorial nationalism. *World Politics*, 66(04):709–746, 2014. 1

Hillel David Soifer. *State Building in Latin America*. Cambridge University Press, 2015. 1

James H Stock and Motohiro Yogo. Testing for weak instruments in linear iv regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, 2005. 6, 7

Sani Umar. The discourses of salafi radicalism and salafi counter-radicalism in nigeria: A case-study of boko haram. *Evanston, IL: Northwestern University*, 156, 2011. 3, 4

Michael Vorenberg. Spielberg's lincoln: The great emancipator returns. *The Journal of the Civil War Era*, 3(4):549–572, 2013. 1

A Appendix

A.1 Proofs

Condition for uniqueness of equilibrium:

Let's define $U(\kappa^*)$ as the left-hand side function of the equation (2). A sufficient condition for a uniqueness of the solution is that the left-hand side increases weakly monotonically in κ^*. Here we follow Morris and Shin [1998, 2004] proof of uniqueness:□

$$U = \kappa^* + \tau\Phi\left(\sqrt{\gamma}(\kappa^* - p)\right), \tag{A.1}$$

We need the derivative of $U(\kappa^*)$ with respect to κ^* to be non-negative:

$$\frac{\partial U}{\partial \kappa^*} = 1 + \tau\sqrt{\gamma}\phi\left(\sqrt{\gamma}(\kappa^* - p)\right) \geq 0, \tag{A.2}$$

$$1 \geq \tau\sqrt{\gamma}\phi\left(\sqrt{\gamma}(\kappa^* - p)\right) \quad \geq \tau\sqrt{\gamma}\tfrac{1}{\sqrt{2\pi}}, \tag{A.3}$$

Here we use the fact that standard normal p.d.f. $(\phi(x))$ maximum value is equal to $\frac{1}{\sqrt{2\pi}}$ at $x = 0$ and substituting $\phi(\cdot)$ with $\frac{1}{\sqrt{2\pi}}$ in equation A.3. This gives the sufficient condition for a unique equilibrium:

$$2\pi \quad \geq \tau^2\gamma \quad \blacksquare \tag{A.4}$$