Uncertainty and Civil War Onset

Iris Malone*

Abstract

Why do some armed groups escalate their campaigns to civil war, while others do not? Only 25% of the 960 armed groups formed between 1970 and 2012 became violent enough to surpass the 25-battle death threshold, often used to demarcate "civil conflict." I develop a new theory that argues this variation occurs because of an information problem. States decide how much counterinsurgency effort to allocate for repression on the basis of observable characteristics about an armed group's initial capabilities, but two scenarios make it harder to get this decision right, increasing the risk of civil war. I use fieldwork interviews with intelligence and defense officials to identify important group characteristics for civil war and apply machine learning methods to test the predictive ability of these indicators. The results show that less visible armed groups in strong states and strong armed groups in weak states are most likely to lead to civil war onset. These findings advance scholarly understanding about why civil wars begin and the effect of uncertainty on conflict.

*Department of Political Science, 100 Encina Hall West, Stanford University, Stanford, CA 94305. (`irismalone@stanford.edu`, `web.stanford.edu/~imalone`).

1 Introduction

Why do some armed groups escalate their campaigns to civil war, while most do not? Using an original dataset, I show that only 25% of the 960 armed groups formed between 1970 and 2012 became violent enough to surpass the 25-battle death threshold, often used to demarcate "civil conflict."[1] Only 8% of armed groups surpassed the higher "civil war" threshold of 1000 fatalities per year. A large amount of this variation is *within* countries – that is, some armed groups escalate and others do not within the same country – and so cannot be explained by country-level characteristics like state strength. I argue that to understand this variation scholars need a theory about the strategy states use to repress armed groups and under what conditions it is effective or not.

States decide how much counterinsurgency effort to allocate for repression on the basis of observable characteristics about an armed group's ex ante capabilities, dedicating more or less effort to match the situation.[2] If a state has relatively good information, then it can dedicate the right amount of effort to suppress the threat. If a state has poor information, then mismatch can occur, increasing the risk of civil war. I show that, paradoxically, this implies the observable characteristics associated with armed group strength will do a poor job predicting which armed groups are at risk for civil war. Instead, characteristics that make an armed group less visible will be more predictive of civil war. Further, this prediction is more likely to hold in strong states than in weak states due to resource constraints. Even if a weak state knows an armed group is strong, it might not be able to repress the group. This theory, developed at greater length in Section 3, predicts that observable indicators of armed group capability will improve forecasts about which groups cause civil war in weak states, but not as much in stronger states.

I test this using insights from fieldwork interviews and applying supervised learning algorithms to a new dataset with detailed information. I interviewed over 30 policy-makers

[1] See Cunningham et al. (2009, 2013) for more on this definition.

[2] Counterinsurgency efforts refer to the combination of sticks and carrots a state may employ to induce an armed group to stop using violence. It is also sometimes called counterterrorism (Hoffman 1998).

from the U.S. intelligence and defense communities about how they identify and respond to emerging insurgent threats. These interviews show that policy-makers gather observable intelligence about an armed group's capabilities, such as its ideology, aims, or ethnic ties, to inform their threat assessments and counterinsurgency responses. In contrast, characteristics associated with more uncertainty about an armed group's capabilities, such as its recent appearance or proximity to a border, make threat assessments noisier and more difficult to act on. I then use cross-validation and random forests to assess how well these different characteristics predict civil war onset.

I show three results consistent with the theory. First, in strong states, only a handful of characteristics predict civil war better than random. Second, in strong states, the relative importance of visibility characteristics is greater compared to observable strength characteristics. Third, observable strength characteristics improve the prediction accuracy of civil war in weak states more than in strong states. The difference in predictive ability is substantively and statistically significant.

This paper contributes to the literature on civil war onset, both theoretically and empirically. Existing work on rebel organization explains how strong armed groups emerge, but not why the state successfully represses some of them, but not others.[3] Other work explains which states are most at risk for civil war, but not which armed groups within these states are most likely to cause them.[4] Together, these theories might imply that strong armed groups in weak states are most likely to lead to civil war, but this cannot explain why some armed groups in strong states lead to civil war or why some armed groups in weak states fall apart. The main barrier to better understanding these strategic interactions has been the lack of data about armed groups that operated below the civil war threshold of violence. To address this gap, this paper introduces a new theory and dataset to better understand how and why civil wars begin.

[3]See, for example, Weinstein (2005), Staniland (2014), Krause (2017), Lewis (2017), and Larson and Lewis (2018)

[4]See, for example, Fearon and Laitin (2003), Collier and Hoeffler (2004), and Hegre and Sambanis (2006)

In the remainder of this paper, I build my case to explain the puzzle. I use insights from the efficient market literature to show how using an armed group's ex ante observable characteristics to shape state efforts leads them to be good predictors of repression, but poor predictors of conflict. Instead, an armed group's propensity for civil war depends on its unobservable characteristics or pure chance. I develop a simple model to illustrate these intuitions and use expert interviews to understand which observable characteristics states rely on. Machine learning techniques are used to assess the accuracy of these predictions. The results show that strong armed groups in weak states and less visible armed groups in strong states are most likely to lead to civil war.

2 Existing Explanations

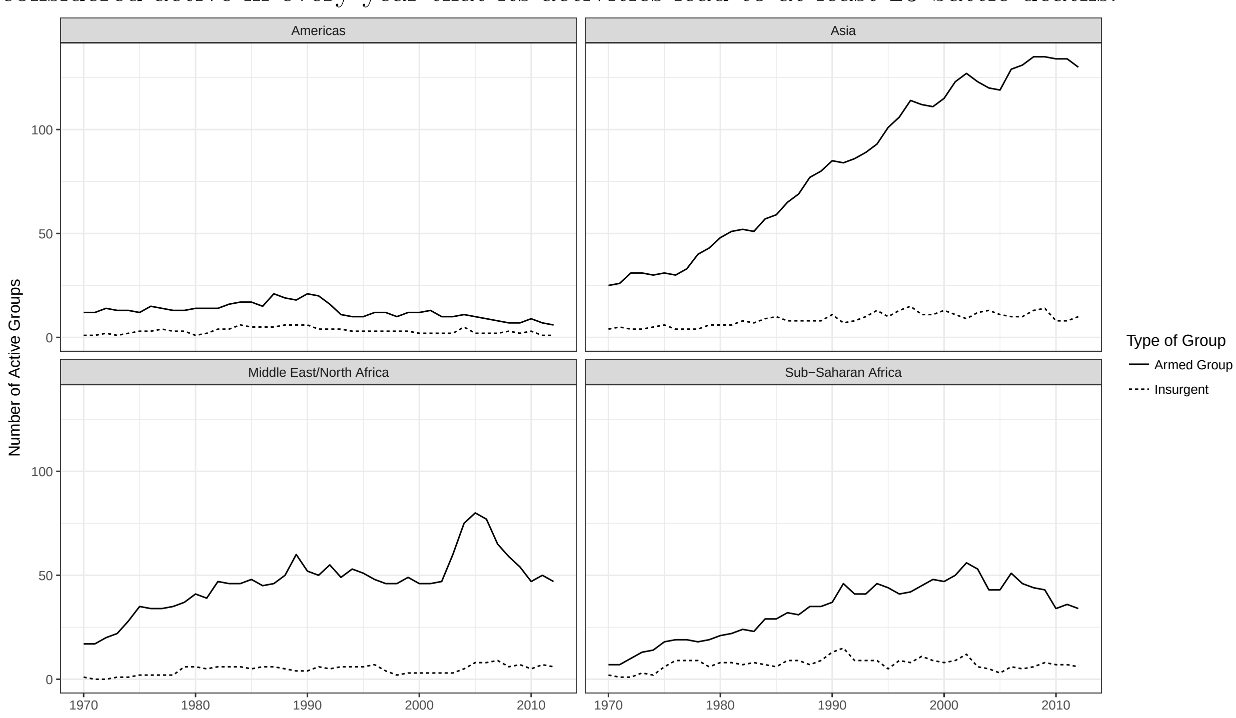
Since the end of the Cold War, civil war has become the most prevalent form of conflict, but few theories have developed to explain how and why civil wars occur. Existing theory focuses on state and rebel capacity, arguing weaker states are more susceptible to civil wars because they cannot accommodate the demands of strong rebel groups. Research on cross-national determinants and in-depth comparative work tends to emphasize relative power between the state and armed group. A central predictor of this approach is that characteristics that increase the relative power of armed groups should explain why some potential insurgent threats become civil wars while others do not.

Cross-national research emphasizes insurgent groups arise in weak states where conditions favor guerrilla warfare and hinder counterinsurgency efforts (Fearon and Laitin 2003). Empirical cross-national analyses find civil war onset most common in states with rough, mountainous terrain, political instability, access to natural resource wealth, and low state capacity (Buhaug and Gates 2002; Hegre and Sambanis 2006).

Inferences about the conditions for civil war based on macro-level data can create an ecological fallacy by generalizing that country-level conditions describe the local conditions where conflicts erupt (Robinson 1950). If these generalizations are inaccurate, then scholars may wrongly attribute certain risk factors to civil war onset that, in reality, have no bearing.

The primary gap in much of this work is that it overlooks how much variation occurs within states. In 30% of civil war onset cases in my data, multiple armed groups were operating in a country at the time of outbreak. In an additional 48% of civil war onset cases in my data, the country was already fighting at least one other active insurgent group. The fact that some armed groups escalate their struggles to civil war while others do not within the same country in the same year suggests characteristics like state strength are insufficient. Figure 1 illustrates the number of active militant organizations, including both armed groups and insurgent groups by region between 1970-2012. It demonstrates just some of the variation between the number of armed groups that launch violent campaigns against the state and the number of those groups that lead to civil war.

Figure 1: **Number of Active Militant Groups by Year, 1970-2012.** These cross-temporal trends come from an original dataset described in Section 4. An armed group is considered active in the year that it first starts using violence. An insurgent group is considered active in every year that its activities lead to at least 25-battle deaths.



State capacity is important, but insufficient to understand civil war onset especially within-state variation. It is necessary to also consider differences across armed group. Comparative research on insurgent groups provides some insight into why certain types of groups might be predisposed to civil war. Stronger armed groups – distinguished by their raw capabilities, strength of grievance, or some combination of the two – are more likely to escalate their campaigns to civil war then weaker armed groups. Motivations to fight a civil war may vary in their intensity across armed groups due to perceived political, economic, or ethnic grievances (Gurr 1970; Horowitz 1985). Armed groups operating in areas where the state discriminates are more likely to become insurgencies because discrimination disproportionately affects certain ethnic groups or social classes and can act as a focal point to spur mobilization within communities. Ethnic groups excluded from power are more likely to fight than ethnic groups, which enjoy some power, because the intensity of their grievance is stronger (Cederman, Wimmer, and Min 2010). Other research argues that "greed" - or an ability to improve an individual's own welfare through rebellion - motivates the rise of insurgencies. In poorer states, individuals have economic incentives to join rebellions if opportunity costs are low and armed groups can pay fighters from the dividends acquired on lootable resources like oil (Collier and Hoeffler 2004; Weinstein 2006).

Some armed groups are able to grow militarily stronger than others. Access to different social networks may also help an armed group accrue enough support to launch an insurgency (Staniland 2014; Lewis and Larson 2018). For example, student or veteran networks provide a natural membership base that can organize quickly before the state detects rebel formation. In the same vein, political aims may also advantage some armed groups over others. Separatist groups may be more likely to become civil wars than center-seeking group because of their natural access to territory and a potential recruitment base (Abrahms 2006; Jones and Libicki 2008; Cronin 2009). Condra (2010) similarly shows that ethnic groups operating further from the capital are more prone to conflict because it is harder for states to control the periphery. These strands of research help explain why some countries are more at risk for civil war and why some armed groups are more adept at insurgent organization. However, the level of analysis can - and should - go even deeper. Generalizing how the state responds to ethnic groups can create another type of ecological fallacy. States respond to armed groups - such as Hamas or Fatah - with different strategies even when they represent the same ethnic group. In order to understand variation across armed groups, scholars need a theory about the strategic interactions between states and armed group.

A limited set of research tries to model these strategic interactions to better explain how and why relative power affects civil war onset. These theories argue that weaker states are more susceptible to civil wars because they cannot accommodate rebel demands. Fearon (2004) identifies, for example, that weak states are susceptible to political or economic shocks that, in turn, temporarily constrain their capabilities further and create a commitment problem. During this shock, the state cannot credibly agree to accommodate the armed group's demands because the armed group knows the state will repress them once the shock dissipates, leading to a bargaining failure and civil war. Civil wars may also occur because power-sharing agreements create a commitment problem between co-conspirators over the future distribution of power. Rulers resolve this through coup-proofing methods that increase their vulnerability to potential insurgent challenges, raising the risk of civil war (Roessler 2016). Others suggest bargaining failures can emerge because uncertainty over a state's resolve can incentivize it to fight armed groups (Walter 2009a, 2009b). These theories suggest civil wars occur due to commitment problems, but these explanations cannot explain why one armed group escalates its campaign to civil war while another does *within* the same country.

An informational theory of war could start to explain this variation. However, scholars often discount information problems as a cause of civil war because of how long they last. At the same time, information problems in civil war settings are not new (c.f. Pye 1964, Taber 1965, Nagl 2002). An armed group has incentives to avoid detection or else risk destruction. This creates uncertainty about an armed group's capabilities making it difficult to pursue effective counterinsurgency strategies.[5] This paper builds on existing counterinsurgency research to propose a new informational theory about how civil wars begin.

3 Theory

I argue civil wars occur, in part, due to asymmetric information about an armed group's capabilities. The state wants to suppress emerging armed groups and allocates counterinsurgency efforts in order to do so. The state decides how much effort to expend against an armed group based on observable characteristics about its' capabilities, dedicating more effort as the severity of the threat increases. However, if a state conditions its efforts on an opponent's observable capabilities, then this information should only predict how much effort the state should expend, not when war will occur (Fearon 1994; Gartzke 1999). Rather, this implies that civil wars occur when the state miscalculates its response against an armed group. I argue miscalculation is more likely against less visible armed groups because they are associated with more uncertainty.

However, this assumes state are relatively unconstrained in their ability to respond to armed groups. In some cases, constrained – or weak states =- cannot mobilize enough resources to suppress an armed group at which point observable characteristics about an armed group start to predict civil war relatively well again. I theorize under what conditions different types of armed groups are more likely to escalate to civil war and derive different testable hypotheses for each case.

This theory is premised on three scope conditions. First, it explains when armed groups lead to new civil conflicts, and not when an armed group joins an ongoing conflict. Second, it primarily explains civil war onset, not civil war recurrence.[6] Third, it focuses on how short-term strategies affect civil war risk, not long-term strategies such as state coup-proofing, political exclusion policies, or rebel organization.

[5]See, for example, Kalyvas (2006), Lyall (2010), Condra and Shapiro (2012), Shapiro and Weidmann (2015), Shapiro et al. (2018)

[6]See Walter (2004) on civil war recurrence.

3.1 Uncertainty and Civil War Onset

A militant threat left unchecked risks evolving into a larger conflict.[7] A state has an incentive to repress an armed group before this happens because an armed group is usually weaker and easier to destroy. The state's strategy to combat an armed group is to allocate time, resources, and manpower to counterinsurgency operations. There is a cost-effectiveness tradeoff in these operations due to the financial, organizational, and political costs involved so the state dedicates more or less effort depending on the situation.

I argue that states decide how much effort to allocate on the basis of observable characteristics about an armed group's expected capabilities. If a state has good information about how strong an armed group is, then it can send an optimal amount of soldiers, weapons, or other resources to suppress the armed group. When a state's repression efforts match the armed group's capabilities, this results in a type of "conflict neutrality"; the probability of civil war becomes independent of an armed group's initial observable characteristics.[8] Paradoxically, this implies that observable indicators of armed group strength will make it very hard to predict which armed groups are at risk for civil war in unconstrained or strong states.

HYPOTHESIS 1: In strong states, indicators of observable strength should poorly predict civil war.

How then can civil war occur? If a state has poor information about an armed group's capabilities, then mismatch can occur, increasing the risk of civil war. Uncertainty makes repressing armed groups much harder to conduct because officials are unsure about the necessary scale and scope of efforts.[9]. Noise negatively affects counterinsurgency planning.

[7] "Once the insurgency achieves a certain character of organization...the conflict reaches the state of civil war" (FM 3-24, 4-1).

[8] The logic here is similar to the efficient market hypothesis in economics, which says stock prices reflect all available information at any given time making it difficult to "beat the market" (Malkiel and Fama 1970; Fama 1991).

[9] In the appendix, I formalize this intuition further to illustrate how these parameters increase the likelihood of civil war. The main comparative static of interest is how varying uncertainty affects the probability of war. In short, the model shows that the probability of war is increasing with respect to the amount of uncertainty surrounding a group and independent of an armed group's capabilities.

If the state underestimates an armed group's strength, then an armed group can continue to grow unabated, leading to civil war, as seen in the cases of the African National Congress in South Africa, Shining Path in Peru, or the Communist Party of Nepal-Maoist in Nepal. If the state overestimates the armed group's strength, then it can – in some cases – effectively destroy the group. However, excessive force also risks harming noncombatants and radicalizing moderates (Crenshaw 1981; Lake 2002; Bueno de Mesquita and Dickinson 2007; Zhukov 2014; Carter 2016). Under this condition, too much force can backfire and lead to an influx of new support as seen by the effect of the Marquetella Massacre (Revolutionary Armed Forces of Colombia, 1964), Operation Blue Star (Khalistan Commando Force, 1984), and Operation Ring (Nagorno-Karabkh, 1991). The state's reaction can lead an armed group to escalate its campaign to civil war soon thereafter.

This logic implies that mismatch occurs when the state is missing information about an armed group. I argue that uncertainty about an armed group is essentially a problem about an armed group's visibility to the state. If the armed group is more visible, then the state has less uncertainty about its' capabilities because the state is better able to collect good information. If the armed group is less visible, then the state is more likely to miscalculate against it. In other words, armed group characteristics associated with being less visible will be more predictive of civil war than characteristics associated with observable strength.

HYPOTHESIS 2: In strong states, characteristics associated with less visibility should be more predictive of civil war than ex ante observable strength characteristics.

One solution to avert potential mismatch, especially underestimation, would be to dedicate extra resources to combat any one armed group. However, this strategy is costly for three reasons. First, spending extra resources against any given threat may be unfeasible due to the organizational or political costs involved. Second, there is an additional opportunity cost to spending resources to combat one group, but not another. When Israel faced Fatah and Hamas in 1987 and Uzbekistan faced Hizb-ut-Tahrir and the Islamic Movement of Uzbekistan in 1999, each state chose to concentrate resources to combat one threat over the other. In so doing, Israel and Uzbekistan inadvertently enabled the other armed group to launch a more formidable challenge. Third, this type of hedging increases the risk of over-allocating resources, which can risk exacerbating the situation through the use of indiscriminate force.

This logic assumes states are relatively unconstrained in their ability to respond to armed groups. However, weak states may lack the same skills or resources to combat armed groups due to budget shocks, corruption, or the need to suppress other armed groups. Even if a weak state knows an armed group is strong, it may not be able to do enough to suppress a strong armed group, increasing the risk of a larger conflict. This predicts that within weak states, observable strength characteristics are more likely to predict which armed groups lead to civil war. Comparing armed groups in weak states to armed groups in strong states should show that characteristics associated with observable strength provide a larger marginal improvement in prediction in weak states over strong states.

HYPOTHESIS 3: Observable strength characteristics should improve the marginal accuracy of civil war better in weak states than in strong states.

Various challenges make it hard to assess the strength of armed groups with any great precision, which consequently increases the likelihood of miscalculation in strong states. In order to test these hypotheses, I turn to expert interviews to better assess what characteristics states rely on to identify observable strength and visibility.

4 Assessing Strength and Visibility

Identifying which indicators states use to assess the observable strength or uncertainty of an armed group requires an expert's knowledge of insurgent threat assessment procedures. I conduct approximately 30 semi-structured interviews with current and former officials from the U.S. National Security Council, Defense Department, and intelligence community. These interviews centered around how officials asses the threat of different armed groups, asking their opinions about how both armed group strength and visibility challenges affect decision-making. Experts often couched their assessments based on insurgent threats that they had the most experience dealing with such as the Philippines or Afghanistan.

When asked what types of armed group characteristics they found to be the most serious indicators of a possible insurgency, officials listed a variety of different features including:

- groups with ethno-nationalist or religious ideologies (Senior Terrorism Expert A; Defense Department Official D)[10]
- groups with ties to certain tribal or ethnic networks (Senior Terrorism Expert C; Intelligence Officer E)[11]
- groups with external logistic support and weaponry (Defense Department Official C; Defense Department Official D; Military Officer C)[12]
- groups with labor or student bases (National Security Council Official A)[13]
- groups with combat experience like ex-police or veteran fighters (Intelligence Officer E; Defense Department Official D)[14]
- groups in areas with conflict legacies (Military Officer D)

[10] "There are huge numbers of untested assumptions early on about [expected strength], but I think here's where some degree of ideology does really matter...we have to make certain assumptions and ideology can play that well" (Interview with Senior Terrorism Expert A 2017)

[11] During a discussion regarding the start of the Iraqi insurgency, one individual noted that a lot of Sunni groups had access to ex-Baathist networks which gave them a pre-existing organizational infrastructure to tap into – this proved rather helpful. It created intelligence networks and an early warning system that then Sunnis could exploit to better avoid detection (Paraphrased Interview with Senior Terrorism Expert C 2017). Similarly, during a discussion regarding the start of the Afghan Taliban in the 1990s, one individual noted fighting in the mujahideen during the 1980s created "extant networks [the Taliban] could activate" to grow stronger. Their identification with different mujahidin militias still resonates 20 years later. "Their communal sense of identity is very powerful." In Pashto, it's called "quam" – the bonds that form through shared experience (Paraphrased Interview with Intelligence Officer E 2017).

[12] "If they get, you know, 'X' type of weaponry, then we're worried." (Interview with Defense Department Official C 2017)

[13] "Young student movements, labor movements - they sometimes have lower opportunity costs to fighting" (Interview with National Security Council Official A 2017)

[14] 'From the 1990s, we knew who was in charge [of the Afghan Taliban] and who to keep tabs on. In 2003, we saw leadership assembling again and thought they're probably going to organize an insurgency.' (Paraphrased Interview with Intelligence Officer E 2017)

Across these interviews, two themes emerged. First, policy-makers rely on a large number of potential characteristics to track armed groups. Political aims, ideology, social networks, ethnic ties, and conflict experience all affect expert assessments about what makes armed groups more serious threats. Second, the indices mentioned most often reflected beliefs about an armed group's observable capabilities. On occasion, other features like resolve, adaptation, or resilience also emerged as important in theory, but hard to measure in practice.

Despite the fact that many experts cited observable indicators to identify potential insurgent groups, many interviewees also recounted their 'surprise' and 'shock' when certain insurgent groups emerged. When I asked officials about why these surprises happen, responses emphasized the unpredictability of conflict and the challenges dealing with large amounts of uncertainty surrounding these groups. As one interviewee remarked "just because we know [a country's] at risk for conflict, that doesn't mean we know exactly what will emerge. In Afghanistan, we could have had more warlords or Al Qaeda or the Taliban. The fact that it was the Taliban doesn't mean we could have predicted that one in advance" (Interview with NSC Official B 2017). Experts repeatedly mentioned the same challenge associated with uncertainty and armed group visibility:

- "When you spend time studying these groups, it's so hard to understand what's going on and the least you can do is be cognizant of this limitation in your decision-making. However, there is so much uncertainty surrounding these groups that people are really biased to focus on the information they already have" (Interview with Defense Department Official D).

- "There is lots of noise at the margins in these assessments which makes it hard to use because people want certainty." (Interview with Intelligence Office C 2017)

- "I think a problem that we have with our intelligence community and use of intelligence is we want prediction. We're trying to get to certainty....And so you magnify that in an insurgency when you have human decision-making and there's no exact science about

> that. To be able to predict the kind of decisions that they're going to make despite not being able to assess with certainty what kind of capabilities they have is difficult." (Interview with Military Officer A 2017)

Uncertainty surrounding an armed group is hard to incorporate into decision-making. As one interviewee put it "you can't base strategy on what might not happen, but what you think will happen" (Interview with Defense Official E 2017). As a result, correcting for uncertainty by investing more resources than a state anticipates needing rarely makes it into the state's counterinsurgency calculations.

If conventional wisdom is correct, then the observable strength characteristics listed in interviews and existing literature should predict civil war onset relatively well. Strong armed groups are defined by their ability to accumulate resources or the intensity of their grievances. Armed groups with certain political aims or ideological goals may be stronger because of their natural access to territory and a potential recruitment base. Groups with combat experience should also be stronger than other types of armed groups. If two existing armed groups merge together, this merger may accumulate capabilities. In contrast, if an armed group forms by splintering away from a parent organization, then this armed group may be weaker. Similarly, certain ethnic armed groups may have more intense ethnic grievances than other armed groups, making them more willing to fight.

The existing literature suggests that these observable characteristics may correlate with more supporters, bases, or willingness to fight and are thus more predictive of civil war. If the theory developed here is correct, then these characteristics should be less likely to predict civil war. Instead, characteristics associated with more uncertainty should be more likely to predict civil war in strong states.

I suggest scholars can measure this uncertainty by considering an armed group's relative visibility. Organizational characteristics that make a group less visible also make it harder for the state to collect good information about an armed group's capabilities.[15] I suggest

[15] For a similar take on legibility and repression, see Scott (1998), Lee (2018), and Blaydes (2018).

there are at least three types of characteristics associated with uncertainty to take into consideration. I also summarize how these different organizational characteristics map onto an armed group's strength and visibility in Table 1.

Table 1: How Organizational Characteristics Correspond to Strength and Uncertainty

Strength	Uncertainty
Political Aims	Rough Terrain
Intense Ethnic Grievance	Transnational Operations
Ideological Goal	Peripheral Operations
Combat Experience	Peripheral Ethnic Groups
Strong Support Base	Recent Appearance

First, armed groups operating in the periphery may reveal less information. Armed groups closer to the border or further from the capital are harder for the state to assess than armed group operating in the center of the country. In these peripheral areas, the state has a harder time monitoring and learning about an armed group's capabilities. Proximity to the border can also make it easier for an armed group to establish an external base, which can provide additional cover and help make it harder to detect the group's strength (Salehyan 2009). Armed groups conducting transnational attacks in other countries may be less visible to the state due to challenges associated with intelligence-sharing (Enders and Sandler 2011). Consistent with previous research about how rough terrain complicates counterinsurgency, I suggest that armed groups operating in inaccessible mountainous terrain are also less visible to the state (Fearon and Laitin 2003).

Second, peripheral ethnic armed groups, which are excluded from power but not actively discriminated against, may be less visible than ethnic armed groups with some political power (Roessler 2016). Irrelevant or powerless groups do not require constant monitoring either in order to accommodate their demands or to actively repress them.

One weakness to the characteristics listed so far is that they are also often associated with an armed group's strength. However, one final variable - recent appearance - goes against this. As an armed group ages, it might have more time to aggregate capabilities, but it also develops a reputation and becomes more visible. Newer armed groups have conducted fewer attacks to signal their potential capabilities, which makes it harder to assess their expected capabilities and incorporate this information into counterinsurgency planning. Although few analyses examine the effect of insurgent age on conflict, similar analysis has found that the age and elapsed tenure of leaders substantially affects conflict onset.[16] Younger leaders and less experienced leaders are more conflict-prone in part due to private information about their resolve (Chiozza and Goemans 2004; Horowitz, McDermott, and Stam 2005; Wolford 2007).

5 Data

A major research obstacle to understanding why some armed groups escalate to civil war has been the lack of cross-national data about armed groups. Existing research on armed groups has generally used extensive fieldwork and in-depth case studies to gain insight on these actors because of the lack of cross-national data. The closest quantitative resource - the Minorities at Risk Organizational Behavior Dataset - only focuses on 112 armed groups in 12 Middle Eastern countries active between 1980 and 2004. Another resource - the Terrorist Organization Profiles dataset - ceased updates in 2008 and was removed from public access.

I address this gap by constructing a new cross-national dataset known as the Armed Group Dataset. The dataset records information about 1,570 armed groups in 72 countries between 1970-2012. For qualitative scholars, I create an encyclopedia of short historical narratives and bibliography for each group totaling over 3,600 pages. For quantitative scholars, I code new variables about the characteristics of armed groups and their violent campaigns.

This dataset builds on the population of armed groups first introduced in the Terrorist Organization (TORG) Crosswalk project (Asal et al. 2015). The crosswalk provides a common identifier framework for different armed groups mentioned in the Global Terrorism Database (GTD), Uppsala Conflict Data Program (UCDP), Terrorist Organization Profiles

[16] One exception is Bapat (2005) on insurgency duration and negotiations.

(TOPS), and Minorities at Risk Organizational Behavior (MAROB) datasets. It also provides estimates about when armed groups started and stopped using violence.[17]

The TORG population records information about 2,750 unique armed groups in 142 countries between 1970 and 2012. Many armed groups in this population are not at-risk for civil war because of the political environments where they emerge. Treating armed groups like Baader-Meinhof (Germany) and Aum Shinrikyo (Japan) as equivalent to the Naqshbandi Army (Iraq) or the Red Sea Afar Democratic Organization (Eritrea) could bias results by drawing inferences about the characteristics favoring civil war across states instead of within them. Consequently, I choose to restrict my analysis to only the 1,570 armed groups active in the 72 countries that experienced at least one civil war between 1970 and 2012.

Proto-insurgency campaigns often launch violent attacks years before a civil war finally erupts.[18] While intuition suggests these campaigns should affect the prospects of a larger conflict, proto-insurgencies are surprisingly absent from current theories about insurgency formation and civil war onset.[19] Political science has historically struggled to identify the population of proto-insurgency campaigns necessary to address this gap. This problem arises because proto-insurgencies have incentives to be clandestine and avoid detection by state forces. Their organizational structures, especially early on, are often amorphous and fluid as seen by the early stages of the insurgency in the Syrian Civil War. While these issues make it harder to identify these groups, it does not make identification impossible. I suggest that political science can identify these groups by considering when and why militant groups use violence.

Terrorism scholars have long characterized the use of political violence by militant groups as strategic because it helps armed groups mobilize resources, support, and legitimacy (Crenshaw 1981; Hoffman 2006; Kydd and Walter 2006). Insurgencies do not form simply when proto-insurgencies have the opportunity to do so. While plenty of militant groups represent

[17] Correspondence with Ken Cousins and Victor Asal. January 2017.

[18] See, for example, violent behavior by the Lord's Resistance Army, Ansar al-Islam, Mujahidin-e-Khalq, or Sendero Luminoso prior to their "official" start dates in common civil war datasets.

[19] See Byman (2008) for an exception.

discriminated populations (e.g. Kurdistan Freedom Hawks), operate in mountainous terrain (e.g. Punjabi Taliban), or have combat experience (e.g. Jemaah Anshorut Tauhid (JAT)), they are still unable to muster enough strength to surprise the state. Insurgencies require popular support, financial backing, and weapons to sustain a violent campaign against a state, but are not all endowed with these resources when they form (Tilly 1978; Lichenbach 1995; Weinstein 2007). Although proto-insurgencies would prefer to remain undetected, the need to amass these capabilities creates incentives to attack police stations, warehouses, or villages for supplies. These early guerrilla attacks force militant groups to reveal their presence early on even if they are not yet able to escalate their campaign. These attacks also enable scholars to identify a population of proto-insurgency campaigns from existing political violence datasets like the Global Terrorism Database.

While media reports often discount political violence as the work of criminals, bandits, or terrorists, these labels are somewhat apocryphal. Many of these militant groups often evolve into credible insurgent organizations through the sustained use of politicized violence against a state.[20] Although some work often refers to these groups as terrorist organizations, it is dangerous to simply discount them as such. First, it invites a long-standing and unresolved definitional debate about what terrorism is which detracts from the overall focus of this project.[21] Second, this paper builds on growing work that suggests insurgent and terrorist organizations should not be studied in isolation from one another, but rather treated as synonymous actors whom all use violence in the pursuit of political ends (Fortna 2015; Asal et al. 2015).

The dataset records new information about the armed group's date of formation, goals, ideology, membership base, ethnic ties, and external support based on different news articles, declassified intelligence assessments, think tank reports, and scholarly sources. There are two possible empirical concerns here. First, a few characteristics are time-varying. Armed

[20]During the 1950s and 1960s, U.S. and British intelligence almost exclusively referred to insurgencies in Malaysia, Vietnam, and Latin America as the work of "Communist Terrorists" or CTs.

[21]See, for example, Schmid and Jongman (1988, 29-38), Enders and Sandler (2006, 3-6), de la Calle and Sanchez-Cuenca (2011).

groups often evolve over time in response to state efforts and aims or areas of operations may change. As a result, I make sure to code information based on an armed group's *initial* violent appearance. This means finding early intelligence assessments or newspaper records about when a group first appeared. It also results in many variables coded as unclear or unstated.

Second, some states have better initial information about an armed group than other states. There may be a concern that researchers know more in hindsight about an armed group than the state initially did. However, the amount of information available about a group is often tied to whether it becomes an insurgent group or not. Since the theory predicts that groups associated with more uncertainty are more likely to become insurgents, the direction of the bias runs counter to what the theory predicts.

The characteristics listed in Table 1 and in the existing literature imply a large number of organizational characteristics could affect both an armed group's strength and lack of visibility. This necessitates researching and coding a large number of variables.

I collect information on the following variables:

- **Political Aim:** A binary variable that denotes the group's initial aims as either territorial or center-seeking (SEPARATIST). If the group's aims are never justified or unclear, this is missing.

- **Ideology:** A series of dichotomous variables indicating the ideological philosophy of the group including leftist, right-wing, ethno-nationalist, religious, anarchist, or environmental. An armed group can have no clear ideology (e.g. Revolutionary United Front) or can have multiple ideologies (e.g. Mujahidin-e-Khalq).

- **Ethnic Network Base:** A series of dichotomous variables indicating the ethnic base of an armed group and its ethnic political status in the year it first used violence based on the Ethnic Power Relations dataset 3.0. The primary variables are DISCRIMINATED, IRRELEVANT, and POWERLESS. The reference category is whether the ethnic base is unclear, the ethnic base has some power, or has autonomous status.[22]

- **Social Network Base:** A series of dichotomous variables indicating the social base of an armed group's leaders or initial members. The primary variables are LABOR, STUDENT, political party or organization POLPARTY, religious organizations or communities (RELIGMEM), and members with prior combat experience including former police, former military, former militants, or foreign fighters (VETERANS). Following on the Global Terrorism Database's coding rules, I code the social base for madrassa recruits as students. The reference category is an unstated membership base.[23]

- **Type of Formation:** A series of dichotomous variables recording how the group formed including whether it is previously unknown (NEW) the result of a splinter (SPLINTER), the result of a merger (MERGER), the result of a political party becoming violent (POLTURNSMIL), or the result of a militia becoming political (MILTURNSPOL).[24]

- **External Activities:** Two binary variables measuring whether an armed group has an external base (EXTERNALBASE) or conducts transnational attacks in other countries (TRANSNATIONAL). The correlation is $\rho = 0.43$.

- **Group Age:** Two continuous variables measuring an armed group's age and recent appearance. AGE is the total length of time an armed group operated between when it first started using violence and its censor date. Armed group age is censored when it stopped being an armed group due to disarmament, repression, or reorganization or first reached the civil war threshold. Observations are also censored in 2012 if an armed group is still operating in that year. INCIPIENT is the length of time an armed group operated between when it first formed and when it first started using violence.

- **Geographic Operations:** A series of continuous variables about where armed groups

[22] I do not separately include an autonomous status variable since it is likely endogenous to the armed group's conflict propensity.

[23] Approximately 49% of armed groups have no clear membership base.

[24] Approximately 61% of armed groups are previously unknown.

operate based on approximately 47,000 geocoded violent attacks listed in the Global Terrorism Database and Global Event Database (Sundberg and Melander 2003).[25] I map an armed group's area of operations based on a 50km buffer zone around where it conducts attacks and incorporate information from the PRIO-GRID spatial dataset about where armed groups operate in the first year they use violence. I measure where armed groups operate, including their logged closest distance to the border (LNBDISTMIN), logged closest distance to the capital (LNCAPDISTMIN) and average mountainous terrain (MTNMEAN) (Tollefsen, Strand, and Buhaug 2012).

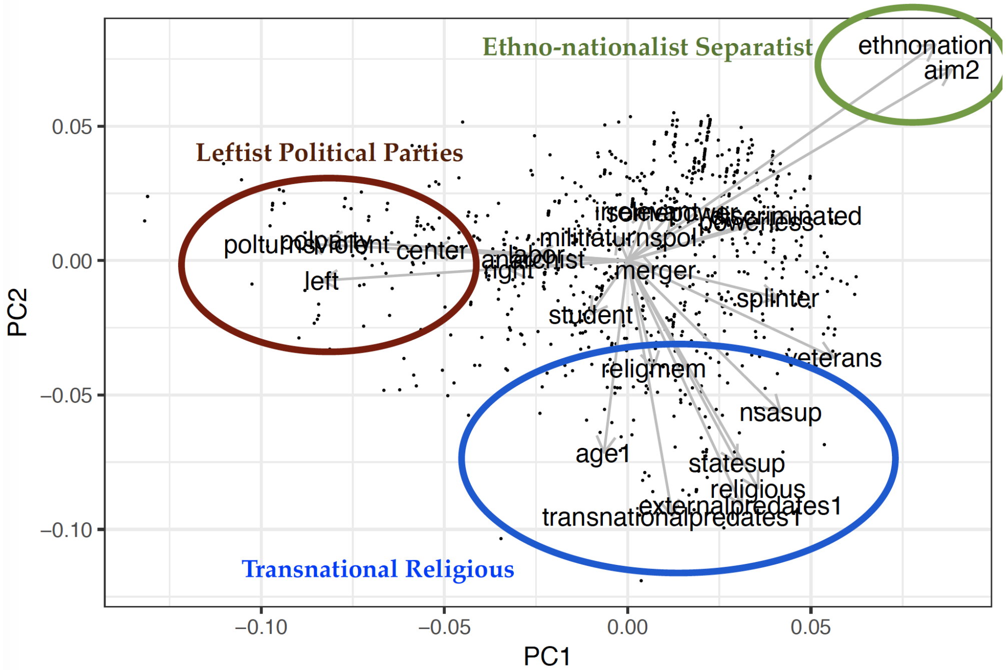
To help visualize the relationship between these variables, Figure 2 presents a principal component analysis of the data. It is an unsupervised learning technique used in machine learning to explore patterns in the data. A principal component analysis shows how different variables correlate with each other when I apply an orthogonal transformation to reduce the dimensionality of the data. It suggests that there are three clusters of variables, which correspond to three different types of armed groups: (1) leftist political parties, e.g. Communist Party of Thailand, People's Revolutionary Party, (2) ethnonationalist separatist groups, e.g. Baloch Liberation Front, Kosovo Liberation Army, and (3) transnational religious groups, e.g. Al Qaeda in Islamic Maghreb, Jemaah Islamiyya.

The three clusters correspond to our historical knowledge about trends in different types of armed groups over time (Figure 2). During the Cold War, communist movements launched guerrilla campaigns in Latin America, Sub-Saharan Africa, and Southeast Asia bolstered in part by Soviet assistance. Decolonization and the fall of the Soviet Union led to a proliferation of new ethnonationalist, separatist campaigns to achieve independence. Finally, the Global Al Qaeda and Global Islamic State networks have grown since 2001, creating a plethora of transnational, Islamist campaigns in the Middle East, North Africa, and Asia.

For the analysis here, I place three additional restrictions on the data such that each

[25] I exclude all attacks by unidentified assailants and insurgent attacks, which occur *after* an armed group reaches the civil war threshold of violence. I also exclude attacks which cannot be geocoded to at least the second-order administrative district.

Figure 2: **Principal Component Analysis of Armed Group Dataset.** This unsupervised learning technique reduces a large number of correlated variables into a smaller number of uncorrelated variables, also known as Principal Components. The first Principal Component on the X-axis seeks to explain as much variation in the data as possible. The second Principal Component on the Y-axis seeks to explain as much of the remaining variation in the data as possible. Each dot represents one armed group in the dataset.



armed group satisfies three criteria: (1) politicized opposition to the state, (2) a unique name identifier, and (3) justified violent activity. The first criteria builds on selection criteria introduced in Staniland (2014) to identify insurgent groups. It omits pro-government militias, criminal gangs, and individual lone wolfs.[26] When the TORG dataset lists an armed group's armed wing and political wing separately, I combine these observations into one record.[27] I also combine records when the group's name is a confirmed alias for another group like the different aliases for Fatah and Lashkar-e-Taiba. These three criteria lead to a sample of 960 armed groups in 72 countries of which 240 escalate to civil wars between 1970-2012.

The unit of analysis in the data is the government-group dyad because many organi-

[26]Less than 1% of pro-government militias in the dataset end up switching sides and becoming politicized. This means that while there is a small risk of omitting some militias, which had the opportunity to turn against the state, their absence is unlikely to substantively affect the findings.

[27]This decision can partially explain why some resources - like the South Asian Terrorist Portal - list more groups in India or Pakistan than I record in the Armed Group Dataset.

zational characteristics are not time-varying. The outcome variable, ONSET, is a binary measure recording whether an armed group reaches a civil war threshold of violence in a particular dyad. It is coded as one if the armed group meets the 25-battle death civil war conflict threshold in the Uppsala Armed Conflict Dataset and zero otherwise (Gleditsch et al. 2002).[28]

A key conditioning variable in the theory is whether a state is strong (unconstrained) or weak (constrained). Conflict is potentially endogenous to economic growth so I identify weak states based on their economic productivity in the decade before my analysis starts. I use data from the World Bank and Penn World Tables to measure a country's GDP per capita, averaging these estimates between 1960-1969 to mitigate against potential outliers. I consider states weak if their logged GDP per capita fall in the bottom quartile of states. If the state formed after 1969, then I use their GDP per capita in the first year of existence. This decision rule leads to a population of 24 weak states, listed fully in the Appendix. I consider the other 48 states in the population strong states.

The dataset introduced here records a large number of variables associated with armed groups. Since it is based on crosswalk data, it is compatible with mergers from other conflict datasets. It promises to not only provide insight into questions about civil war, but other important research questions in terrorism and insurgency studies.

6 Results

I use different supervised learning techniques to assess the theory. The convention in quantitative conflict research is to rely on parametric regression methods for hypothesis-testing. However, these regression methods are not the best suited to answer the question posed here due to the large number of variables there are to consider. Including all of the variables listed in Table 1 runs up against two possible problems. First, I do not know which of these vari-

[28]See Lacina (2006) for more on battle deaths and Uppsala's measurement. This measure requires both state and non-state actors to clash. It is different from Uppsala's one-sided violence measure where an armed group or state unilaterally kills 25 state or non-state actors in an attack.

ables are important; including irrelevant variables can increase the size of the standard error, producing inefficient estimates (Angrist and Pischke 2009, 64). Second, these variables can create a dimensionality problem because the number of predictors is very large compared to the number of observations. This runs the risk of model overfitting and producing spurious correlations (Gill 1999).

I build on a growing set of research that gets around these issues by using machine learning techniques to examine the correlates of conflict.[29] My hypotheses are all centered around predicting civil war onset – a task at which machine learning excels. Prediction requires the creation of an out-of-sample validation set to test how well a model built around one set of data does on another set. I accomplish this by randomly partitioning the data into a training and test (validation) set for each state type. The training set for strong states contains 573 groups and the test set contains 207 groups.[30] The training set for weak states contains 148 groups and the test set contains 90 groups.

Since the outcome variable, ONSET, is binary, I build a classifier. The main goal of classification is to minimize the prediction error – or misclassification rate – on a training set and maximize the predictive accuracy of the model on a test set. Conflict occurs in about 20% of the strong state sample and is relatively balanced across both the training and test sets. Conflict occurs in about 42% of the weak state sample and is also relatively balanced across both the training and test sets. However, as a robustness check in the Appendix, I oversample the strong state subset results to 35% and show it does not affect the findings. Table 2 lists the variables I put into the classifier.

Methods Selection

I test the theory's hypotheses using two algorithms: k-fold cross-validation for a Classification and Regression Tree (CART) and Random Forest (RF). These algorithms require very few

[29] See, for example, Hill and Jones (2014), Chiba and Gleditsch (2017), Colaresi and Mahmood (2017), Cederman and Weidmann (2018), Montgomery and Olivella (2018), and Mueller and Rauh (2018).

[30] Transnational groups appear in multiple countries.

Table 2: Summary of Variables in Classifier

Splinter (SPLINTER)	Anarchist (ANARCHIST)	Student Base (STUDENTS)
Merger (MERGER)	Log Minimum Distance to Border (LNBDISTMIN)	Veterans Base (VETERANS)
Militia Turns Political (MILITIATURNSPOL)	Log Minimum Distance to Capital (LNCAPDISTMIN)	Labor Base (LABOR)
Political Party Turns Violent (POLTURNSVIOL)	Average Percent of Mountainous Terrain (MTNMEAN)	Religious Base (RELIGMEM)
Leftist (LEFT)	Ethnic Group Status - Powerless (POWERLESS)	Political Party (POLPARTY)
Religious (RELIGIOUS)	Ethnic Group Status - Discriminated (DISCRIMINATED)	External Base (EXTERNALBASE)
Right-Wing (RIGHT)	Ethnic Group Status - Irrelevant (IRRELEVANT)	Transnational Attacks (TRANSNATIONAL)
Ethnonationalist (ETHONATIONALIST)	Total Campaign Duration (LNAGE)	
Separatist (SEPARATIST)	Time Between Forming and Using Violence (INCIPIENT)	

assumptions about the data and provide a lot of model flexibility in case there are important nonlinear relationships in the data. CART is a nonparametric decision tree. The algorithm creates a set of decision rules, choosing how to split the data at various nodes in order to best explain variation in the outcome variable.

Cross-validation is a well-established method to ensure the model chooses decision rules that best minimize the prediction error (Efron 1983; Friedman et al. 2013). K-fold cross-validation divides the data into k subsets, trains the model on every subset except an i^{th} fold, and then computes the test error on the i^{th} set. The cross-validation rate is the number of k subsets that produces the lowest training error on the dataset. In most cases, the lowest training error in cross-validation also results in the lowest validation set error. I use nested cross-validation techniques to "prune" the tree. The logic behind pruning is to remove irrelevant predictors that do not substantially change the model's prediction error. It identifies irrelevant predictors by creating a decision tree based on all the variables and comparing it to a permutation of a decision tree that remove that variable. If the classification error substantially increases with that variable's exclusion, then the variable is deemed important. If the classification error does not change substantially, it is removed. The final result is a more parsimonious model without sacrificing any predictive accuracy (Friedman et al. 2013, 308). It reduces the variance of the model without adding much bias.

A cross-validated CART is a simple and easy way to visualize the relationship, but their high variance can result in poor prediction rates on out-of-sample data. It relies on a "greedy" algorithm, meaning it chooses decision rules based on what reduces the local mean squared error rather than the global mean squared error. Partitioning the training and test set in different ways can also result in different decision rules, further reducing its interpretability. The use of a more sophisticated classification algorithm known as a random forest can overcome these limits.

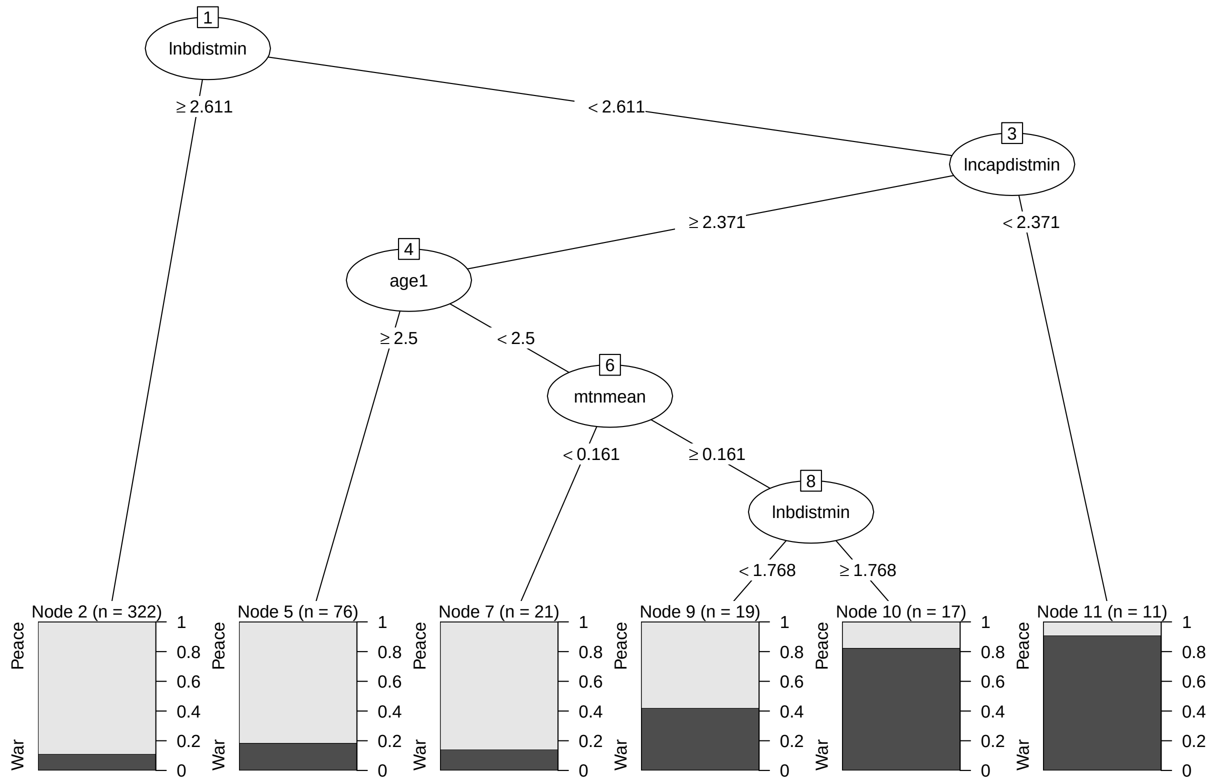
Random forests bootstrap different observations B times from the training set and construct a single CART for each bootstrapped sample. Each time the tree considers a new decision rule, it does so by considering a split among 6 random variables chosen from Table 2. This helps decorrelate the predictions between trees. It creates a different tree B times, then averages the predictions from across the different trees to identify the relative importance of each variable. This reduces the variance problem inherent in one decision tree. I now turn to test the theory using these methods.

Hypothesis 1 and 2

The first two hypotheses both consider what types of armed groups cause civil war onset in strong states. The first hypothesis predicts that few, if any, observable strength characteristics should be predictive of civil war onset in strong states. The second hypothesis predicts that characteristics associated with less visibility should be relatively more important than observable strength characteristics in strong states. I use cross-validation and random forests to assess these claims.[31] Figure 3 shows the results of a 6-fold pruned CART using cross-validation.

[31] Additional cross-validation tuning information is available in the Appendix.

Figure 3: **6-Fold Cross-Validated Tree for Strong States.** The training set results of a pruned decision tree using 6-fold cross-validation. Each node represents a new decision rule. All additional observation under that branch have that organizational characteristic. The nodes at the bottom show the proportion of armed groups correctly classified under each of those rules as well as the number of observations associated with that rule.



The tree visualizes the different decision rules node by node. Each node takes one variable that explains a lot of variation between armed groups that escalate and those do not. It then creates a decision rule – or proposed cut in the data to partition the armed groups. Variables that are listed further towards the top explain relatively more variation in the outcome variables than variables further down. For example, all observations under the left branch of LNBDISTMIN (node 1) are armed groups in strong states operating more than 14 km ($e^{2.61}$) away from the border. Armed groups operating less than 14 km away from the border in strong states are on the right side of the split. The tree iteratively makes decision rules that best explain variation. The bins at the bottom of the tree capture how many armed groups actually surpass the civil conflict threshold ("war") and those that do not ("peace"). For example, approximately 10% of armed groups operating more than 14 km away from the border escalate to civil war (node 2). In contrast, 20% of armed groups that are less than 14 km away from the border (node 1), more than 11km away from the capital (node 3), and have been operating for more than 2.5 years (node 4) escalate to civil war.

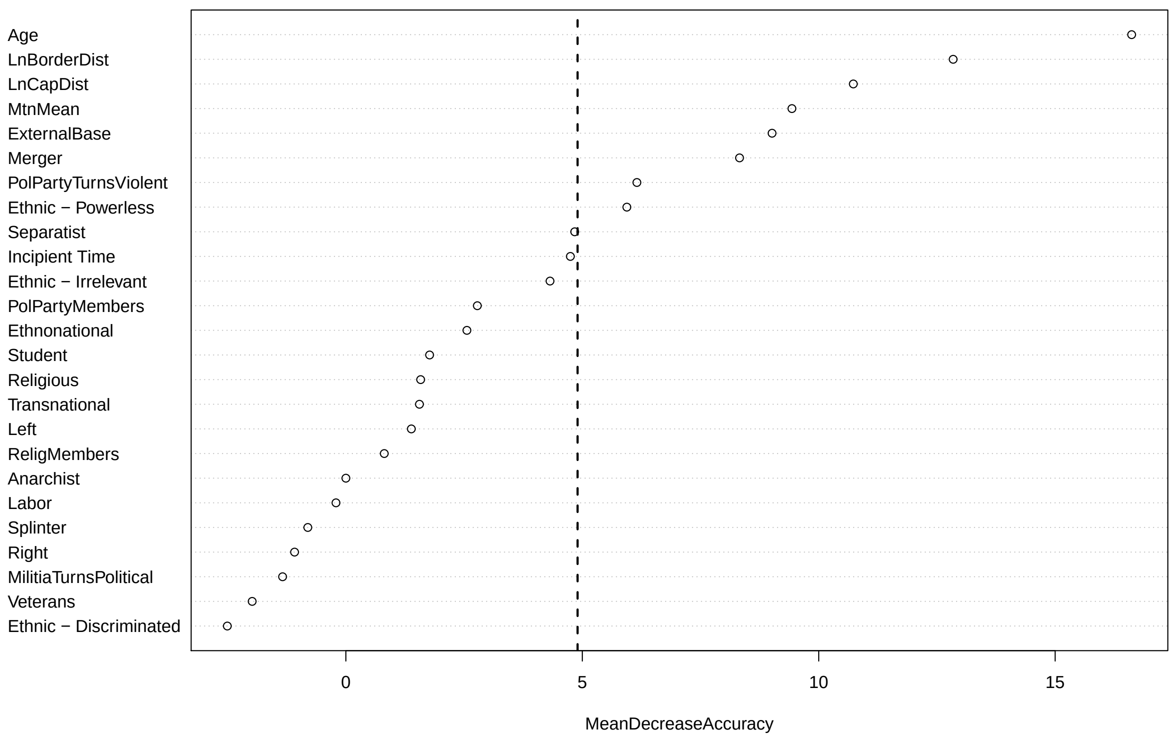
The decision tree provides two pieces of evidence to support the theory. First, I find that nested cross-validation techniques remove 84% of the input variables, shrinking from 25 to 4 characteristics. Most of the observable strength variables are pruned off the tree because they do not substantially affect the misclassification rate. This is consistent with the first hypothesis that relatively few variables should actually predict civil war in strong states if policy-makers incorporate information about these ex ante characteristics into counterinsurgency planning.

Second, the variables that do remain on the tree and are relatively important predictors are characteristics associated with more uncertainty. Distance, group age, and mountainous terrain are all features that reduce the state's ability to get good information about a group's capabilities. Notably, political aims, ideology, and social networks are absent from the pruned model. This does not mean that they have no explanatory power, but rather their explanatory power is relatively limited compared to a group's age or border distance.

I also build a random forest to provide a more challenging test of these hypotheses. The standard way to illustrate the results of a random forest is through a variable importance plot (Figure 4). In random forests, the relative importance of a variable is measured by how much it results in a *mean decrease in accuracy*. The mean decrease in accuracy metric is straightforward. It measures how much the global predictive accuracy of the model changes when a permutation of the model excludes that one variable. The more a variable reduces the accuracy of the model, the more important it is.

I create a measure to compare whether this mean decrease in accuracy is significant by creating a placebo check. I include a random uniform model in the classifier and run the classifier to see how much this placebo variable reduces the mean decrease in accuracy. The dashed line in Figure 4 is the mean decrease in accuracy for the placebo. It also provides a threshold to identify which variables are more important and fall to the right of the line.

Figure 4: **Relative Variable Importance Plot for Strong States.** The dashed line is the relative importance of a random uniform variable added to the model as a placebo check. If a variable does better than this placebo, it is considered significant and falls to the right of the line. If a variable does worse than this placebo, it is considered insignificant and falls to the left of the line.



The random forest results provide further support for Hypothesis 1 and 2. The importance plot suggests that 68% of the input variables are insignificant predictors, consistent with Hypothesis 1. Further, the types of variables deemed relatively more important are the characteristics associated with more uncertainty. Of the 8 variables identified as relatively important, the first five are visibility-based characteristics and the last three are observable strength characteristics. The relative importance of visibility characteristics is higher than the observable strength characteristics, backing Hypothesis 2.

Hypothesis 3

The final hypothesis compares the predictive accuracy of civil war onset between strong and weak states. If the third hypothesis is correct, then incorporating all available information about an armed group's observable strength should lead to a larger marginal improvement in weak states than in strong states. In order to draw comparisons between these two samples, I build a random forest using all 25 variables listed in Table 2 and run it for both types of states. The variable importance plot for the weak state subset is listed in the Appendix.

A receiver operating curve (ROC) illustrate basic differences in classification accuracy across the two samples. ROC plots the sensitivity classification rate against the specificity classification rate.[32] The Area Under the Curve (AUC) measures how well the classifier correctly predicts whether an armed group leads to civil war or not. If a classifier does better, the AUC is larger. Figure 5 show that the AUC is larger for weak states than it is for strong states.

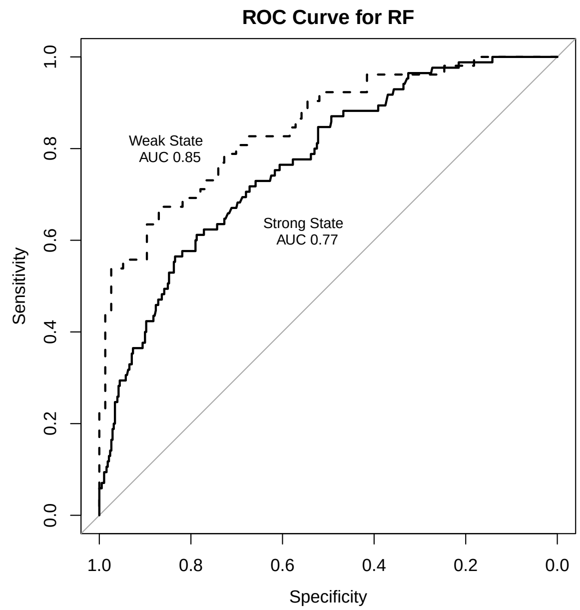
A decision tree with no information is one where it just predicts that an armed group fails to reach the civil war threshold every single time. The no information rate in weak states is 58%; the no information rate in strong states is 81%. Since there are potential class imbalance issues, I run a secondary set of cases where I oversample armed groups in the strong state subset until the two samples have the same classification rate of 50%. I tune the random forest to bootstrap 500 different decision trees for each sample and thus end up with a sample of the predictive accuracy of 1000 individual trees.[33]

I compare the out-of-bag error estimates across the two samples. The out-of-bag error measures the error rate when this model is applied to an out-of-bag (validation) set of observations. When the number of trees is sufficiently large, the out-of-bag error is similar

[32] Recall that sensitivity also known as the true positive rate measures the number of true positives over the sum total of true positives and false negatives. Specificity, also known as the true negative rate, measures the number of true negatives over the sum of true negatives and false positives.

[33] Additional random forest tuning information is available in the Appendix.

Figure 5: **ROC Curve for Strong State and Weak States.** The area under the curve measures the model's accuracy rate. Values closer to one represent a better classification model.

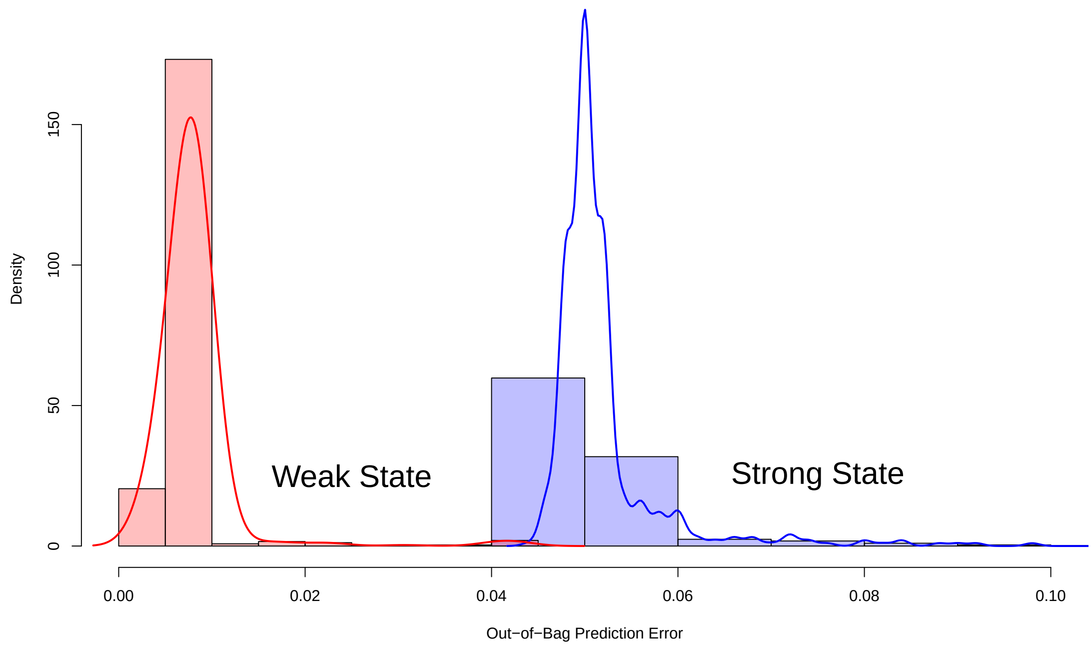


to leave one out cross-validation. The accuracy is one minus the out-of-bag error on the training set. The histogram below show the out-of-bag prediction error rates from the different bootstrapped trees.

Figure 6 illustrates that there are large differences in the out-of-bag error rates for armed groups in weak states versus strong states. While the distribution for strong states is nearly centered around 0.05, the distribution for weak states is centered around 0.01. Given the same amount of information in each model, the error rate is very small for the weak state sample. By contrast, the out-of-bag error rate for the strong state sample is much larger. This implies that different observable information predicts civil war onset among armed groups in the weak state subset better than the strong state subset consistent with the third hypothesis.

I also compare the F_1 statistic from the result across the two samples. The F_1 statistic is sometimes preferable for imbalanced samples. In these situations, the out-of-bag error or prediction accuracy are relatively uninformative metrics because the number of true negative

Figure 6: Histogram of Out-of-Bag Prediction Error



observations is large. The F_1 statistic records the average precision and recall rates to assess the model's performance. It is useful when the costs of a false positive are different than the costs of a false negative. Since civil conflict is often costlier in the long-run than wasting some resources against one armed group, comparing the precision and recall rates may be preferable. Larger F_1 statistics suggest the model does a better job explaining cases of war and vice versa. The F_1 statistic for the weak state sample is 0.689 and the F_1 statistic for the strong state sample is 0.381.

A limit to decision trees is that they can only predict which variables are important. They cannot tell scholars much about the direction or the magnitude of a variable's relationship to conflict. Further, as a decision tree grows larger (e.g. has more branches), it loses generalizability because it is overfitting to the training data. A linear probability model with a few well-justified variables can test whether a variable is substantively and statistically significant. As a final way to assess the third hypothesis, I run a linear probability model for the strong and weak state samples, arbitrarily including the top five variables based on their mean decrease in accuracy score. The model is run with and without country fixed effects to control for time-invariant factors within countries that could affect an armed group's propensity to escalate. Standard errors are clustered by country to control for heterogeneity. The results are shown in Table 3.

Table 3: **Cross-Sectional Model of Civil War Onset for Weak and Strong States.** This is a linear probability model measuring whether an armed group leads to civil war (ONSET). The outcome is whether an armed group causes a civil war that leads to at least 25-battle deaths. The variables in this model are chosen because a Random Forest algorithm identifies these characteristics as the five most important variables for each subset.

Subset	Model 1 Weak	Model 2 Weak	Model 3 Strong	Model 4 Strong
DISCRIMINATED	0.54^{***}	0.48^{***}		
	(0.10)	(0.06)		
MERGER	0.35^{***}	0.32^{***}		
	(0.10)	(0.10)		
LNAGE	-0.08	-0.09^{*}	-0.05^{***}	-0.04^{***}
	(0.05)	(0.05)	(0.02)	(0.02)
LNCAPDISTMIN	0.03	0.01	0.01	0.01
	(0.03)	(0.03)	(0.01)	(0.01)
MTNMEAN	-0.34^{***}	-0.10	0.12	0.21^{***}
	(0.11)	(0.14)	(0.07)	(0.08)
LNBDISTMIN			-0.05^{***}	-0.05^{***}
			(0.01)	(0.01)
EXTERNALBASE			0.16^{***}	0.10^{**}
			(0.04)	(0.04)
Country FE	N	Y	N	Y
R^2	0.33	0.49	0.09	0.23
Adj. R^2	0.31	0.42	0.08	0.16
Num. obs.	143	143	545	545
RMSE	0.41	0.38	0.38	0.36

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors clustered by country.

The results provide additional insight into how different organizational characteristics change the probability of civil war onset. In all these models, LNAGE is negative and – in most models – statistically significant. In Model 2, for example, a 1% increase in an armed group's age corresponds to a 9 percentage point decrease in the probability an armed group in a weak state becomes embroiled in civil war, holding all other variables constant. In Model 4, a 1% increase in an armed group's age corresponds to a 4 percentage point decrease in the probability an armed group in a strong state leads to civil war, holding all other variables constant. Similarly, two observable strength characteristics in weak states, DISCRIMINATED and MERGER, are positive and statistically significant. Increasing armed group strength is positively associated with civil war onset. Conversely, characteristics associated with less visibility in strong states, like LNBDISTMIN and EXTERNALBASE, increase the probability of civil war onset.

In the Appendix, I test these hypotheses using alternate machine learning tuning parameters and stricter higher civil war thresholds. The results do not change substantially, providing support for the theory.

7 Conclusion

This paper advanced a new theory to explain why some armed groups escalate their campaigns to civil war while others do not. In strong states, efforts to repress less visible armed groups sometimes fail because of miscalculation. In weak states, efforts to repress strong armed groups sometimes fail because of resource constraints.

Using an original dataset and machine learning techniques, I find evidence to support this theory. Because the state conditions its efforts based on observable strength characteristics, very few of these indicators predict civil war onset. Less visible armed groups - characterized by age, distance, and area of operations - are more likely to cause civil wars in strong states. Strong armed groups - characterized by ethnic ties, organizational structures, or political aims - are more likely to cause civil wars in weak states.

For policymakers, the research provides insight into conflict prevention efforts. The growing fragility of countries in the Sahel, Middle East, and Southeast Asia has led to a proliferation of violent extremism and armed groups. However, existing research on civil war risk typically compares threats across countries, overlooking massive variation within states. If states better understood which armed groups in their countries are most serious, they could better develop targeted strategies to de-escalate conflicts early on.

For scholars, this research advances academic debates about the causes of civil war and effect of uncertainty on war. It provides a more fine-grained analysis of why some armed groups lead to civil war, but other do not. This paper also introduces opportunities for research on other pathways to civil war. Domestic political incentives, for example, also likely affect the strategies states use to fight armed groups. There is a rich set of research to be done on the strategic interactions underpinning civil conflict.

References

[1] Abrahms, Max. "Why Terrorism Does Not Work." *International Security* 31, no. 2 (2006): 42-78.

[2] Angrist, Joshua and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press: Princeton. 2009.

[3] Asal, Victor, Ken Cousins, and Kristian Skrede Gleditsch. "Making Ends Meet Combining Organizational Data in Contentious Politics." *Journal of Peace Research* 52, no. 1 (2015): 134-138.

[4] Bapat, Navin A. "Insurgency and the Opening of Peace Processes." *Journal of Peace Research* 42, no. 6 (2005): 699-717.

[5] Berman, Eli, Joseph H. Felter, and Jacob N. Shapiro. *Small Wars, Big Data: The Information Revolution in Modern Conflict.* Princeton University Press. 2018.

[6] Blainey, Geoffrey. *Causes of war.* Simon and Schuster. 1988.

[7] Blaydes, Lisa. *State of Repression: Iraq Under Saddam Hussein.* Princeton University Press, 2018.

[8] Bueno de Mesquita, Ethan, and Eric S. Dickson. "The propaganda of the deed: Terrorism, counterterrorism, and mobilization." *American Journal of Political Science.* 51, no. 2 (2007): 364-381.

[9] Buhaug, Halvard, and Scott Gates. "The Geography of Civil War." *Journal of Peace Research* 39, no. 4 (2002): 417-433.

[10] Buhaug, Halvard, and Jan Ketil Rod. "Local Determinants of African Civil Wars, 1970-2001." *Political Geography* 25, no. 3 (2006): 315-335.

[11] Byman, Daniel. "Understanding Proto-Insurgencies." *The Journal of Strategic Studies* 31, no. 2 (2008): 165-200.

[12] Carter, David B. "Provocation and the strategy of terrorist and guerrilla attacks." *International Organization.* 70, no. 1 (2016): 133-173.

[13] Cederman, Lars-Erik, Andreas Wimmer, and Brian Min. "Why Do Ethnic Groups Rebel? New Data and Analysis." *World Politics* 62, no. 1 (2010): 87-119.

[14] Cederman, Lars-Erik, and Nils B. Weidmann. "Predicting armed conflict: Time to adjust our expectations?." Science 355, no. 6324 (2017): 474-476.

[15] Chiozza, Giacomo, and Hein E. Goemans. "International Conflict and the Tenure of Leaders: Is War Still Ex Post Inefficient?" *American Journal of Political Science* 48, no. 3 (2004): 604-619.

[16] Colaresi, Michael, and Zuhaib Mahmood. "Do the robot: Lessons from machine learning to improve conflict forecasting." *Journal of Peace Research.* 54, no. 2 (2017): 193-214.

[17] Collier, Paul, and Anke Hoeffler. "Greed and Grievance in Civil War." *Oxford Economic Papers* 56, no. 4 (2004): 563-595.

[18] Condra, Luke Nayef. "Ethnic Group Rebellion in Civil War." PhD diss., Stanford University, 2010.

[19] Condra, Luke N., and Jacob N. Shapiro. "Who takes the blame? The strategic effects of collateral damage." *American Journal of Political Science.* 56, no. 1 (2012): 167-187.

[20] Crenshaw, Martha. "The Causes of Terrorism." *Comparative Politics* 13, no. 4 (1981): 379-399.

[21] Cronin, Audrey Kurth. *How Terrorism Ends: Understanding the Decline and Demise of Terrorist Campaigns.* Princeton University Press, 2009.

[22] Efron, Bradley. "Estimating the error rate of a prediction rule: improvement on cross-validation." Journal of the American statistical association 78, no. 382 (1983): 316-331.

[23] Enders, Walter and Todd Sandler. *Political Economy of Terrorism.* Cambridge University Press. 2011.

[24] Fama, Eugene F. "Efficient capital markets: II." *The Journal of Finance* 46, no. 5 (1991): 1575-1617.

[25] Fearon, James D. "Signaling versus the balance of power and interests: An empirical test of a crisis bargaining model." Journal of Conflict Resolution 38, no. 2 (1994): 236-269.

[26] —. "Rationalist Explanations for War." *International Organization* 49 (1995): 379.

[27] —. "Why Do Some Civil Wars Last so Much Longer than Others?" *Journal of Peace Research* 41, no. 3 (2004): 275-301.

[28] Fearon, James D., and David D. Laitin. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97, no. 1 (2003): 75-90.

[29] "Field Manual 3-24: Counterinsurgency." Department of the Army, Washington DC. 2006.

[30] Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning.* Springer Series: New York. 2013.

[31] Gartzke, Erik. “War is in the Error Term.” *International Organization* 53, no. 3 (1999): 567-587.

[32] Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Haavard Strand. “Armed Conflict 1946-2001: A New Dataset.” *Journal of Peace Research* 39, no. 5 (2002): 615-637.

[33] Gurr, Ted Robert. *Why Men Rebel.* Routledge, 1970.

[34] Hegre, Havard, and Nicholas Sambanis. “Sensitivity Analysis of Empirical Results on Civil War Onset.” *Journal of Conflict Resolution* 50, no. 4 (2006): 508-535.

[35] Hill, Daniel W., and Zachary M. Jones. “An empirical evaluation of explanations for state repression.” *American Political Science Review* 108, no. 3 (2014): 661-687.

[36] Hoffman, Bruce. *Inside Terrorism.* Columbia University Press, 2006.

[37] Horowitz, Donald L. *Ethnic Groups in Conflict.* University of California Press, 1985.

[38] Horowitz, Michael C., Rose McDermott, and Allan C. Stam. “Leader Age, Regime Type, and Violent International Relations.” *Journal of Conflict Resolution.* 49, no. 5 (2005): 661-685.

[39] Kalyvas, Stathis N. *The Logic of Violence in Civil War.* Cambridge: Cambridge Press. 2006.

[40] Krause, Peter. *Rebel Power: Why National Movements Compete, Fight, and Win.* Cornell University Press, 2017.

[41] Lacina, Bethany. “Explaining the severity of civil wars.” *Journal of Conflict Resolution.* 50, no. 2 (2006): 276-289.

[42] Lake, David A. “Rational Extremism: Understanding Terrorism in the Twenty-First Century.” *Dialogue IO* 1, no. 1 (2002): 15-28.

[43] Larson, Jennifer M., and Janet I. Lewis. “Ethnic Grievances or Ethnic Networks? The Influence of Rumor Networks on Rebel Group Formation.” *International Organization.* 2018. Forthcoming.

[44] Lee, Melissa M. “The international politics of incomplete sovereignty: How hostile neighbors weaken the state.” *International Organization* 72, no. 2 (2018): 283-315.

[45] Lewis, Janet. "How Does Ethnic Rebellion Start?" *Comparative Political Studies.* 2017.

[46] Lyall, Jason. "Are coethnics more effective counterinsurgents? Evidence from the second Chechen war." *American Political Science Review.* 104, no. 1 (2010): 1-20.

[47] Malkiel, Burton G., and Eugene F. Fama. "Efficient capital markets: A review of theory and empirical work." *The Journal of Finance* 25, no. 2 (1970): 383-417.

[48] Montgomery, Jacob M., and Santiago Olivella. "Tree-Based Models for Political Science Data." American Journal of Political Science (2018).

[49] Mueller, Hannes, and Christopher Rauh. "Reading Between the Lines: Prediction of Political Violence Using Newspaper Text." *American Political Science Review* 112, no. 2 (2018): 358-375.

[50] National Consortium for the Study of Terrorism and Responses to Terrorism (START). *Global Terrorism Database [Data File]*. 2017 ed., `https://www.start.umd.edu/gtd`

[51] Nagl, John A. *Counterinsurgency lessons from Malaya and Vietnam: learning to eat soup with a knife.* Westport, CT: Praeger, 2002.

[52] Pye, Lucien. "The Roots of Insurgency." *Internal War: Problems and Approaches.* Ed. Harry Eckstein. Free Press: New York. 1964.

[53] Robinson William. "Ecological correlations and the behavior of individuals." *American Sociological Review.* 15 (1950): 35157

[54] Roessler, Philip. *Ethnic Politics and State Power in Africa: The Logic of the Coup-Civil War Trap.* Cambridge University Press, 2016.

[55] Salehyan, Idean. *Rebels without Borders: Transnational Insurgencies in World Politics.* Cornell University Press, 2009.

[56] Salehyan, Idean, Kristian Skrede Gleditsch, and David E. Cunningham. "Explaining External Support for Insurgent Groups." *International Organization* 65, no. 4 (2011): 709-744.

[57] Scott, James C. *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed.* Yale University Press, 1998.

[58] Shapiro, Jacob N., and Nils B. Weidmann. "Is the phone mightier than the sword? Cellphones and insurgent violence in Iraq." *International Organization.* 69, no. 2 (2015): 247-274.

[59] Staniland, Paul. *Networks of Rebellion: Explaining Insurgent Cohesion and Collapse.* Cornell University Press, 2014.

[60] Sundberg, Ralph, and Erik Melander. "Introducing the UCDP georeferenced event dataset." *Journal of Peace Research* 50, no. 4 (2013): 523-532.

[61] Taber, Robert. *The war of the flea.* Lyle Stuart, 1965.

[62] Tollefsen, Andreas Foro, Havard Strand, and Halvard Buhaug. "PRIO-GRID: A unified spatial data structure." *Journal of Peace Research* 49, no. 2 (2012): 363-374.

[63] Walter, Barbara F. "Does conflict beget conflict? Explaining recurring civil war." *Journal of Peace Research* 41, no. 3 (2004): 371-388.

[64] —. "Bargaining Failures and Civil War." *Annual Review of Political Science* 12 (2009a): 243-261.

[65] —. *Reputation and Civil War: Why Separatist Conflicts Are so Violent.* Cambridge University Press, 2009b.

[66] Weinstein, Jeremy M. *Inside Rebellion: The Politics of Insurgent Violence.* Cambridge University Press, 2006.

[67] Wolford, Scott. "The Turnover Trap: New Leaders, Reputation, and International Conflict." *American Journal of Political Science* 51, no. 4 (2007): 772-788.

[68] Zhukov, Yuri M. "Theory of Indiscriminate Violence." Ph.D. Dissertation. Harvard University. 2014.

8 Appendix

The Appendix contains the following information:

- Appendix A: Formal Model of Theory
- Appendix B: Armed Group Dataset Descriptive Statistics
- Appendix C: Supplemental Machine Learning Specifications
- Appendix D: Robustness Checks
- Appendix E: Sample Armed Group Dataset Narratives

8.1 Appendix A: Formal Model of Theory.

Formal Model

I formalize the strategic logic behind this theory as a decision problem with two periods and two players - the state government G and a potential rebel group R.[34] In the first period, G decides how much counterinsurgency effort $c \in [0, B]$ to expend against R, where B is a budget constraint. Strong states have larger budgets than weak states. In the second period, stochastic elements determine whether R accumulates enough capabilities to surpass a violence threshold $T > 0$, demarcating civil war onset. The rebel group's second period capabilities are $x \in R$ and are drawn from the cumulative distribution function $F(x)$ with mean $\mu - c$ and known variance σ^2. From the government's perspective in the first period, the probability that R surpasses the threshold is $1 - F(T; \mu - c, \sigma^2)$. The state's initial beliefs about R are summarized by μ and σ. The mean, μ, represents the state's initial threat assessment about R's expected capabilities based on its ex ante observable strength characteristics. The variance, σ^2, is the state's confidence around this strength estimate. These distribution represents the ex ante probability an armed group would lead to civil war if the state did nothing against it. However, the state expects it can weaken an armed group's expected capabilities μ by allocating more counterinsurgency efforts c against R. These efforts shift an armed group's capabilities downward.

Suppose the payoff for G if R's campaign stays below the threshold T is $v > 0$ and zero if R causes a civil war. As a result, v measures the political stakes of fighting R relative to the opportunity cost of spending from B. The decision problem for G is to choose how much effort c to expend in order to maximize: $U(c) = vPr(z < T) + (B - c) = vF(T; \mu - c) + (B - c)$

If $F(c)$ is from a location-scale family, such as a Normal distribution, then this can be rewritten as $vF(T - \mu - c) + (B - c)$ where F has mean zero and variance σ^2. The results hold for other F, but a normal distribution yields an analytic solution. The Proofs are in

[34] I thank James Fearon for his large assistance in the set-up and analysis of this model.

the Appendix.

A first-order question is how varying the state's beliefs affects the level of counterinsurgency effort it expends:

Proposition 1. If G's assessment of R is highly uncertain (σ^2 is sufficiently large), then G's optimal effort is $c^* = 0$ regardless of its beliefs about μ.

It is inefficient for G to expend any effort against R when it is highly uncertain about R's expected capabilities. The likelihood that the state's counterinsurgency efforts work is small. If R's expected capabilities are drawn from a distribution with a small variance, then most of the time its actual capabilities will be close to the mean. If R's expected capabilities are drawn from a distribution with a high variance, then it is more likely its actual capabilities will be either smaller or larger than μ. If G believes it is facing a moderately strong R, but in reality ends up facing a weak R ($x << \mu$), then counterinsurgency efforts are not necessary, and thus inefficient. If G believes it is facing a moderately strong R, but in reality ends up facing a very strong R ($x >> \mu$), then counterinsurgency efforts can be insufficient to weaken R and again be inefficient. In other words, when σ^2 is sufficiently large, the effect of varying COIN effort c on the probability that R surpasses the threshold is very low, and not worth the effort. The marginal benefit of additional counterinsurgency effort is very low.

Proposition 2. If G's optimal effort c^* is interior (e.g. $c^* \in [0, B]$) , then the probability R gets over the threshold in period 2 is independent of μ and dependent on σ.

G conditions its effort against R based on observable characteristics about R's expected capabilities, dedicating more or less to match the situation. Since c^* is a function of μ and the probability of war depends on $T - \mu - c^*$, substituting the optimal c into the probability of war causes μ to drop out of the expression for the probability function. Instead, the probability of war depends on σ^2. If G believes it is facing R with expected capabilities μ, but has poor information, then there is a risk G does not dedicate enough effort to suppress the threat. This leads to a higher probability in period 2 that R is strong enough to cause a civil war compared to if G had good information and less uncertainty. This finding should seem counterintuitive in civil war settings. If an armed group's strength does not matter, then why do strong armed groups seem to overrun weak states so often? The next proposition reconciles why:

Proposition 3. If G's optimal effort c^* is at the upper bound of B (e.g. $c^* = B$), then the probability R gets over the threshold in period 2 is dependent on μ when σ^2 is sufficiently small.

In weak states, resource constraints can lead to $c^* > B$. When this happens, G can only allocate B, raising the risk that G's response is insufficient to repress a group even when G is highly confident about the threat it faces. The probability of war is increasing with respect to μ. This increases the parameter space for war in weak states relative to the parameter space for war in strong states at lower levels of uncertainty. It also matches intuition. Civil wars are far more common in weak states than in strong states because the probability G's response is inadequate and R can surpass the threshold is so much higher.

Overallocation Problem.

The above intuition only explains cases of under-allocation. Uncertainty causes the state to invest less effort against less visible armed groups because the marginal payoff to any unit of counterinsurgency effort is smaller. I next consider the second case where over-allocation can occur.

If G wants to mitigate the risk of under-allocation, then it could hedge and allocate some level of effort beyond the optimal level c^*. These efforts can lower the probability R gets over the threshold in the second period, but it can also carry increase the probability of civil war. Why? Up until this point, the model assumed the probability R could get over the threshold was consistently decreasing in c. However, c can vary in its effectiveness. This is because additional levels of effort may weaken R, but can also have a negative externality on the noncombatant population where R operates. Excessive effort against R risks radicalizing moderates or harming noncombatants, particularly if G has uncertainty about who is a combatant and who is a non-combatant when it launches its efforts.

The probability G's efforts are effective can be written as a function of how much effort G expends. Counterinsurgency effort is effective in degrading R's capabilities with probability $p \in [0, 1]$ and ineffective with probability $1 - p$. The probability p captures the amount of precision G has in its effort with higher levels of precision corresponding to more effective efforts. In other words, c is more effective as efforts more precisely target combatants.Under this condition, the decision problem for G is to choose how much effort c to expend in order to maximize: $U(c) = vF(T; \mu - (pc - (1 - p)c) + (B - c)$.

The level of counterinsurgency effort a state expends shifts as a result of its expected effectiveness.

Proposition 4. If G is largely certain about its assessment(σ small), then G's optimal effort is zero if $p < 0.5$ and increasing in p otherwise.

The optimal amount of effort shifts as a function of σ and p. If p is very small, then it is inefficient to expend any effort against an armed group even if the state's assessment about R is fairly certain because the state's effort will not have any appreciable effect. In other words, when the probability of effectiveness is less than 50%, the marginal costs associated with G's exceed the marginal returns from said effort.

Proposition 5. If G is largely uncertain about its assessment(σ large), then G's optimal effort is increasing in p if $p < 0.5$ and zero otherwise.

If the probability of effectiveness is very high, then we recover the same results as Proposition 1 and 2. However, if p is very small *and* the state's assessment about R is fairly uncertain, then the potential gains of weakening R are worth the additional costs. In these circumstances, G can benefit from hedging and allocating additional effort to combat R.

Proof

Assume a Normal distribution where the probability of civil war in period 1 is $F(x) \sim N(\mu, \sigma^2)$ and the probability of civil war in period 2 is $F(x) \sim N(\mu - c, \sigma^2)$:

$$U(x) = vPr(z < T) + (B - c) = vF(T; \mu - c) + (B - c)$$

The first-order condition with respect to c is:

$$\begin{aligned}
0 &= \frac{v}{\sigma}\phi\Big(\frac{T - (\mu - c)}{\sigma}\Big) - 1 \\
\frac{\sigma}{v} &= \phi\Big(\frac{T - \mu + c}{\sigma}\Big) \\
\frac{\sigma}{v} &= \frac{1}{\sqrt{2\pi\sigma^2}}exp\Big(\frac{-(T - \mu + c)^2}{2\sigma^2}\Big) \\
c^* &= \mu - T - \sqrt{2\sigma^2 log\Big(\frac{\sigma^2\sqrt{2\pi}}{v}\Big)}
\end{aligned}$$

Proof of Proposition 1.

The optimal amount of effort is increasing in μ and decreasing in σ. The limit of c^* as σ grows arbitrarily large is negative. Since it is impossible for a state to spend negative effort, it follows that its optimal effort at high levels of uncertainty is $c^* = 0$.

Proof of Proposition 2.

The main comparative static is how the probability of civil war in the second period varies as a function of μ and σ. The probability R is strong enough to launch a civil war is given by the cumulative distribution function $F(x) \sim N(\mu - c, \sigma^2)$. The probability of civil war is $1 - Pr(z < T)$. Substituting in the state's optimal c^* to this function, μ drops out because the state allocates effort based on its beliefs about μ. Consequently, the probability of war varies on the basis of σ.

Proof of Proposition 3.

Weak states are constrained by $c^* \in [0, B]$ such that:

(1) If $c^* \geq B$, then the state allocates $c^* = B$

(2) If $c^* < B$, then the state allocates $c^* = \mu - T - \sqrt{2\sigma^2 log\Big(\frac{\sigma^2\sqrt{2\pi}}{v}\Big)}$

As σ grows arbitrarily small, c^* depends more and more on μ. If the state is increasingly constrained (e.g. B is shrinking), then the probability of war is increasing in μ because the state is more likely to underallocate resources to combat it.

Proof of Proposition 4.
The state prefers to hedge and expend additional c as the effectiveness of these efforts increases. Solving $U(c^*)$ for c finds the set of best responses for the state as p changes. The limit of c^* when σ grows arbitrarily small is $\frac{\mu-T}{2p-1}$. When efforts are completely ineffective ($p = 0$), the optimal level of effort is negative. Since it is impossible for a state to spend negative effort, it follows that its optimal effort at low levels of ineffectiveness is $c^* = 0$. When efforts are more effective than ineffective ($p > 0.5$), the optimal level of effort is positive and increasing in p.

Proof of Proposition 5.
The limit of c^* when σ grows arbitrarily large is negative. If $p < 0.5$, then the quotient is positive and c^* is arbitrarily large and increasing in p. If $p < 0.5$, then the quotient is negative and it is inefficient for G to expend anything.

8.2 Appendix B: Descriptive Statistics

Table 4: Armed Group Dataset

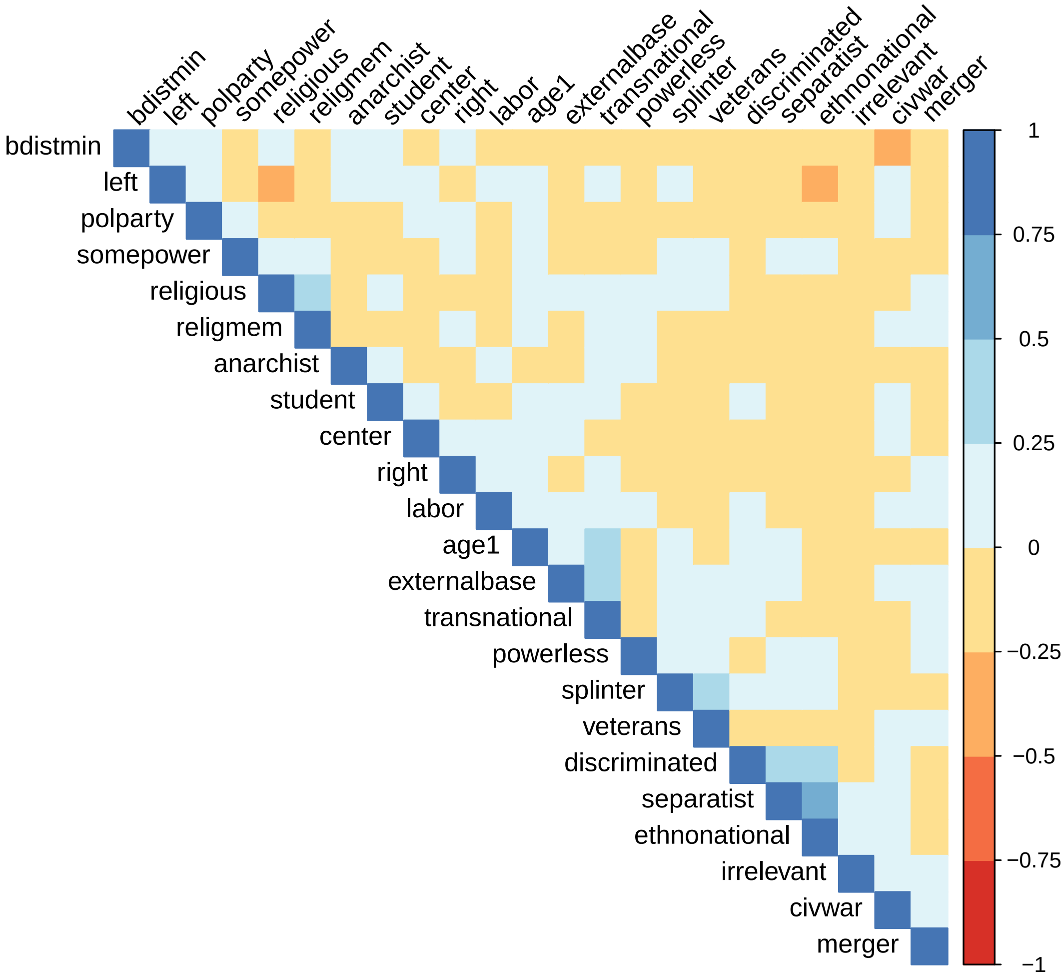
	Statistic	N	Mean	St. Dev.	Min	Max
Dependent Variable	Onset	1,018	0.26	0.44	0	1
	Onset1000	1,017	0.08	0.28	0	1
Political Aim	Separatist	951	0.37	0.48	1	2
Ideology	Left	1,018	0.17	0.37	0	1
	Center	1,018	0.02	0.14	0	1
	Right	1,018	0.03	0.17	0	1
	Religious	1,018	0.30	0.46	0	1
	Ethno-national	1,018	0.48	0.50	0	1
	Anarchist	1,017	0.01	0.08	0	1
	Environmental	1,017	0.01	0.08	0	1
Age	Age Since First Violent	1,011	5.69	8.73	0	65
	Age Since Formation	1,017	8.08	11.40	0	84
EPR Status	Discriminated	1,018	0.12	0.32	0	1
	Powerless	1,018	0.13	0.34	0	1
	Irrelevant	1,018	0.06	0.23	0	1
	Autonomous	1,018	0.04	0.21	0	1
	Some Power	1,018	0.16	0.37	0	1
Social Base	Labor	1,018	0.03	0.17	0	1
	Student	1,018	0.06	0.24	0	1
	Veterans	1,018	0.29	0.45	0	1
	Political Party	1,018	0.10	0.30	0	1
	Religious Organization	1,018	0.03	0.16	0	1
Type Form	New	1,018	0.61	0.49	0	1
	Merger	1,018	0.06	0.24	0	1
	Splinter	1,018	0.25	0.43	0	1
	Party Turns Violent	1,018	0.07	0.25	0	1
	Militia Turns Political	1,018	0.02	0.13	0	1
Transnational	External Base	1,018	0.22	0.41	0	1
	State Support	1,016	0.30	0.46	0	1
	Non-State Actor Support	1,017	0.45	0.50	0	1

Table 5: Armed Group Dataset: Insurgents Only

	Statistic	N	Mean	St. Dev.	Min	Max
Dependent Variable	Onset	260	1.00	0.00	1	1
	Onset1000	259	0.33	0.47	0	1
Political Aim	Separatist	260	0.42	0.50	0	1
Ideology	Left	260	0.22	0.42	0	1
	Center	260	0.03	0.16	0	1
	Right	260	0.02	0.12	0	1
	Religious	260	0.27	0.45	0	1
	Ethno-National	259	0.58	0.49	0	1
	Anarchist	260	0.00	0.00	0	0
	Environmental	260	0.01	0.09	0	1
Age	Age Since First Violent	260	3.76	6.65	0	40
	Age Since Formation	260	5.87	8.90	0	69
EPR Status	Discriminated	260	0.20	0.40	0	1
	Powerless	260	0.15	0.36	0	1
	Irrelevant	260	0.09	0.29	0	1
	Autonomous	260	0.09	0.29	0	1
	Some Power	260	0.15	0.36	0	1
Social Base	Labor	260	0.04	0.20	0	1
	Student	260	0.08	0.28	0	1
	Veterans	260	0.36	0.48	0	1
	Political Party	260	0.13	0.34	0	1
	Religious Organization	260	0.04	0.19	0	1
Type Form	New	260	0.50	0.50	0	1
	Merger	260	0.14	0.35	0	1
	Splinter	260	0.26	0.44	0	1
	Party Turns Violent	260	0.09	0.29	0	1
	Militia Turns Political	260	0.01	0.11	0	1
Transnational	External Base	260	0.28	0.45	0	1
	State Support	260	0.58	0.49	0	1
	Non-State Support	260	0.56	0.50	0	1

N.b. 260 armed groups in the dataset eventually meet the 25-battle death threshold. However, only 240 armed groups reach this threshold between 1970-2012.

Figure 7: Correlation Matrix of Armed Group Dataset



List of Weak and Strong States

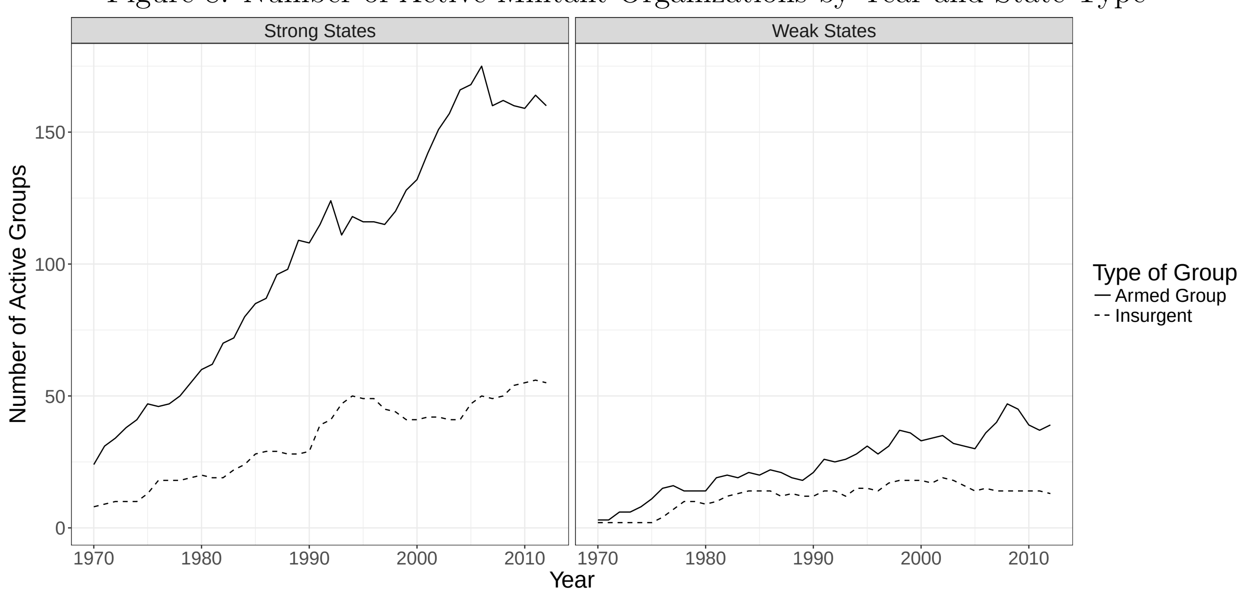
The main measure to identify a weak state is whether it's GDP per capita falls in the bottom quartile between 1960-1969.

List of 'Weak' States based on GDP Quartile: Mali, Mauritania, Chad, Uganda, Burundi, Rwanda, Somalia, Ethiopia, Yemen, Nepal, Congo, Mozambique, Egypt, China, Bosnia-Herzegovina, Sierra Leone, Central African Republic, Democratic Republic of Congo, Sudan, El Salvador, Eritrea, Burma, Indonesia, South Sudan

As a robustness check, I also create a sample of weak states based on whether the UN considered the state a Least Developed Country in 1971.

The list of 'Weak' States based on UN LDC is[35]: Haiti, Mali, Liberia, Mauritania, Chad, Uganda, Burundi, Rwanda, Somalia, Ethiopia, Mozambique, Yemen, Bangladesh, Nepal, Cambodia, Niger, Sierra Leone, Central African Republic, Democratic Republic of Congo, Comoros, Sudan, Afghanistan, Burma

Figure 8: Number of Active Militant Organizations by Year and State Type

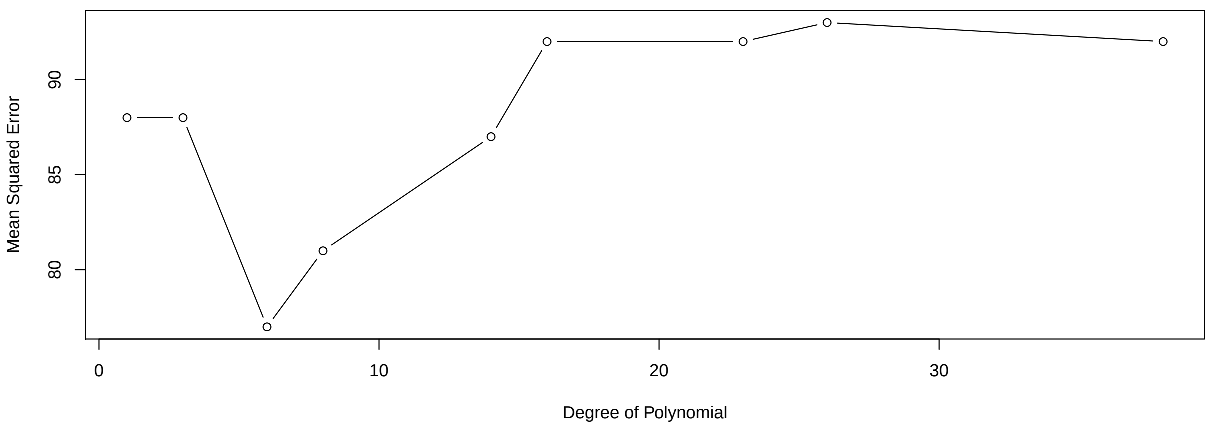


[35] https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/ldc_list.pdf

Appendix C: Supplemental Machine Learning Specifications

Cross-Validation Tuning
The cross-validation rate for the strong state sample is chosen based on the number of folds that minimizes the mean squared error. The plot showing how the mean squared error varies for a large number of folds is shown below.

Figure 9: **Cross-Validation Rate for Strong State Sample.** This plot shows the test error that results from predicting civil war based on varying the number of folds of the dataset. The results show that the minimum test error is 6-fold followed by 8-fold, 14-fold, and 3-fold cross validation.



For comparison, I show the classification and regression trees for different k-fold cross-validation rates. The prediction accuracy of a 8-fold CV CART is 74.4% and a 3-fold CV CART is 76.8%. The prediction accuracy of the higher cross-validation rates is lower than the 76.8% prediction accuracy of a 6-fold CV CART. Simpler models do better than more complicated models.

Figure 10: **8-Fold CV CART for Strong State Sample.** This shows the decision tree for an 8-fold CV CART for the strong sample. .

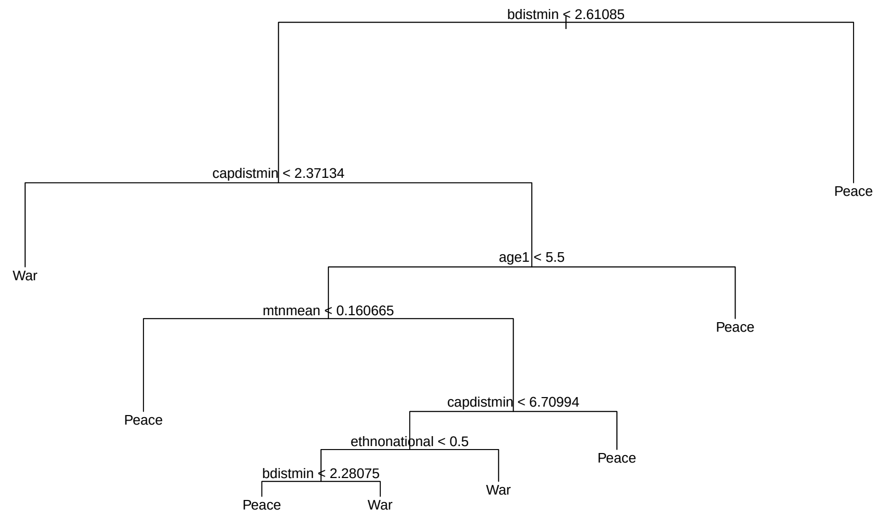
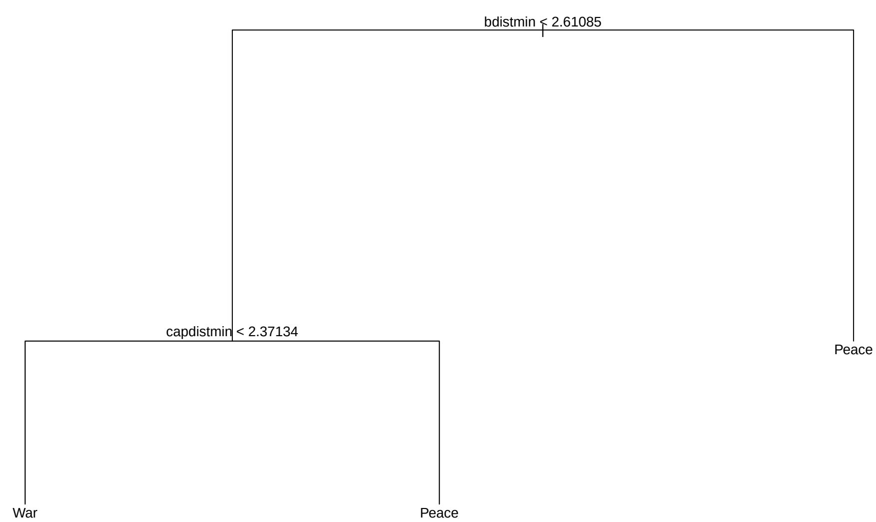


Figure 11: **3-Fold CV CART for Strong State Sample.** This shows the decision tree for an 3-fold CV CART for the strong sample. .



Random Forest Tuning

A random forest is a collection of B different decision trees. Since a single tree is susceptible to high variance, adding additional trees and averaging across their out-of-bag error estimates can result in a lower overall training error. For each random forest, I choose to sample $m = \sqrt{p}$ predictors, which is standard. As an alternate, I try $m = p$ predictors and find a worse error estimate. In order to tune the model, I examine the out-of-bag error by the number of trees for each sample. The results show that as the number of trees grows larger, the average out-of-bag error rate converges towards some minimum value. I choose 500 trees for the random forest. The out-of-bag error estimate for the weak state sample shows a lot of variance, which is expected given the relatively few number of observations.

Figure 12: **Out-of-Bag Error by Number of Trees for Weak State Sample.** This shows the random forest out-of-bag error rate for the weak sample as the number of trees in the random forest grows.

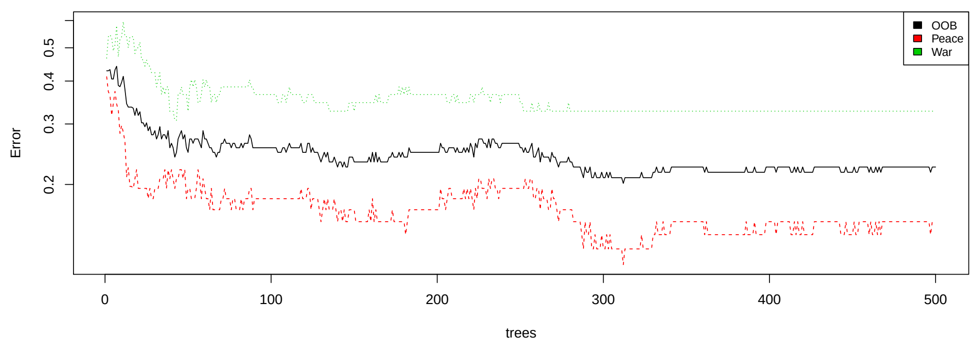
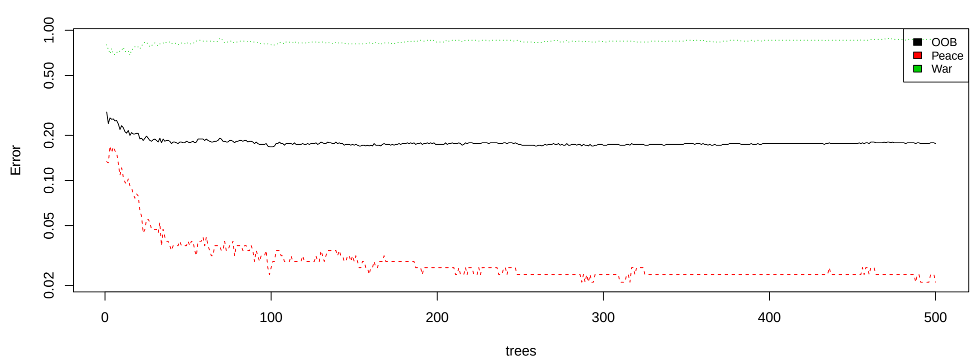
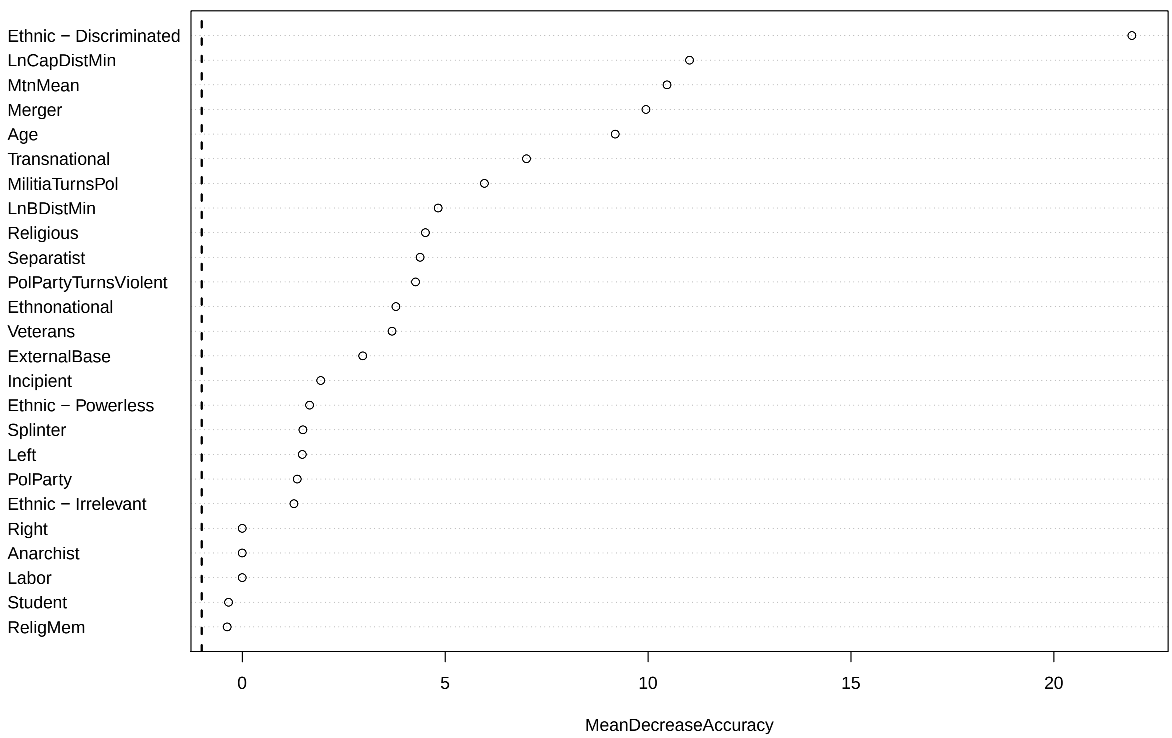


Figure 13: **Out-of-Bag Error by Number of Trees for Strong State Sample.** This shows the random forest out-of-bag error rate for the strong sample as the number of trees in the random forest grows.



For comparison to Figure 4, I also include a relative variable importance plot for the weak states sample. The results show the difference in the relative importance of very variables. Consistent with the theory, Figure 18 shows that observable strength characteristics are considered large predictors. The results show that relative to a placebo check, all of the variables put in the classifier algorithm are considered significant predictors of civil war onset in weak states. The results also show that the variables which are considered more important are those associated with more observable strength characteristics.

Figure 14: **Relative Variable Importance Plot for Weak States.** The dotted line is the relative importance of a random uniform variable added to the model as a placebo check. If a variable does better than this placebo, it is considered significant and falls to the right of the line. If a variable does worse than this placebo, it is considered insignificant and falls to the left of the line.

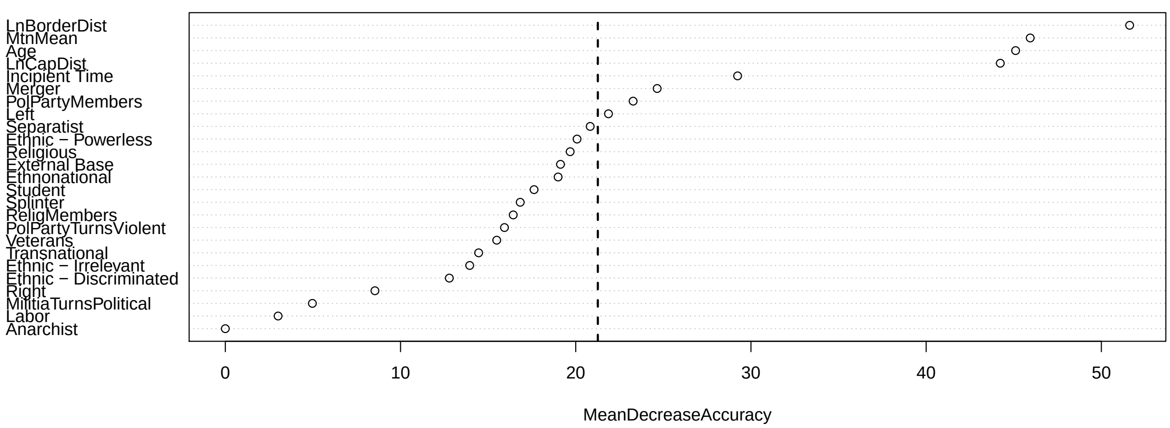


8.3 Appendix D: Robustness Checks

Alternate Specification: Oversample Civil War for Strong State Subset

Since conflict is rare, there is a risk of imbalance. Classifier algorithms do not do well when the outcome variable is very rare and produce very high misclassification errors. One method to correct for this is oversampling and synthetic balance sampling. I use oversample for the strong state sample based on an ROC comparison between oversampling and synthetic balancing. I sample to increase the occurrence of civil war to 35% in the training sample. The results of a Random Forest are shown below. The findings show that the relatively more important variables - age, border distance, capital distance, and mountainous terrain - are all variables associated with an armed group's visibility.

Figure 15: **Relative Variable Importance Plot for Strong States.** The dotted line is the relative importance of a random uniform variable added to the model as a placebo check. If a variable does better than this placebo, it is considered significant and falls to the right of the line. If a variable does worse than this placebo, it is considered insignificant and falls to the left of the line.



Alternate Specification: Higher Civil War Threshold

I run the results using a high civil war threshold of at least 1000 battle deaths per year. I use a 1000-battle death threshold instead of the 25-battle death threshold. Some scholars use a higher battle death threshold to exclude smaller conflicts like the EZLN uprising in Mexico or al-Gama'at al-Islamiyya's conflict in Egypt. It is possible that less visible armed groups can reach the 25-battle death threshold, but only strong armed groups have the means to sustain a lethal campaign, capable of reaching the 1000-battle death threshold. Using a different threshold barely changes the results. Few variables are relatively important predictors of civil war in strong states. Border distance, age, and mountainous terrain are still correlated with civil war onset in strong states.

I use synthetic balance sampling for the weak state sampling based on an ROC comparison between oversampling and synthetic balancing. I use oversampling for the strong state sample based on an ROC comparison between oversampling and synthetic balancing. I sample to increase the occurrence of civil war to 25% in the training sample. The results of these different samples is shown below.

Hypothesis 1 and 2

I provide the variable selection results from a cross-validated tree and random forest for the strong state sample.

Figure 16: **6-Fold Cross-Validated Tree for Strong States.** The training set results of a pruned decision tree using 6-fold cross-validation. Each node represents a new decision rule. All additional observation under that branch have that organizational characteristic. The nodes at the bottom show the proportion of armed groups correctly classified under each of those rules as well as the number of observations associated with that rule.

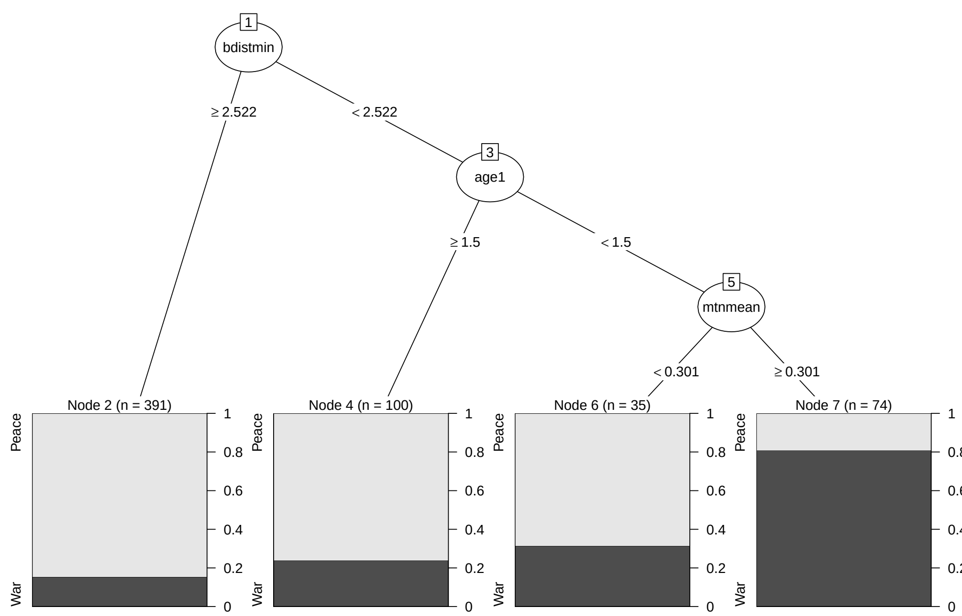
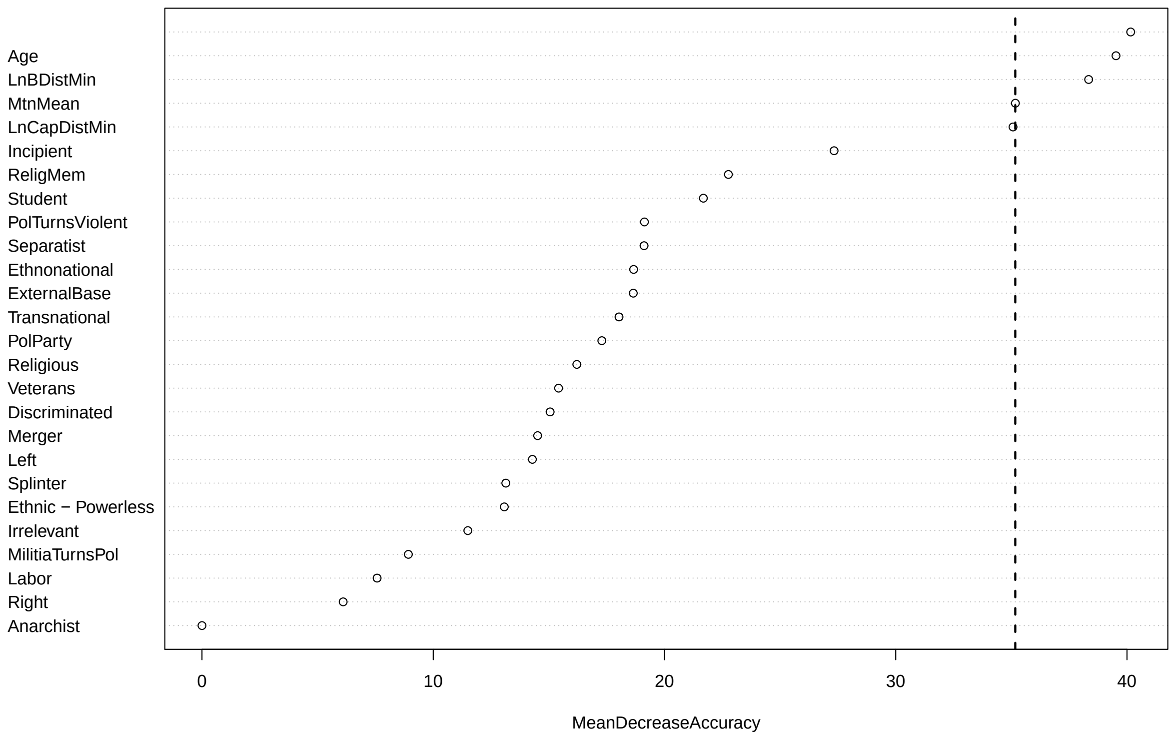
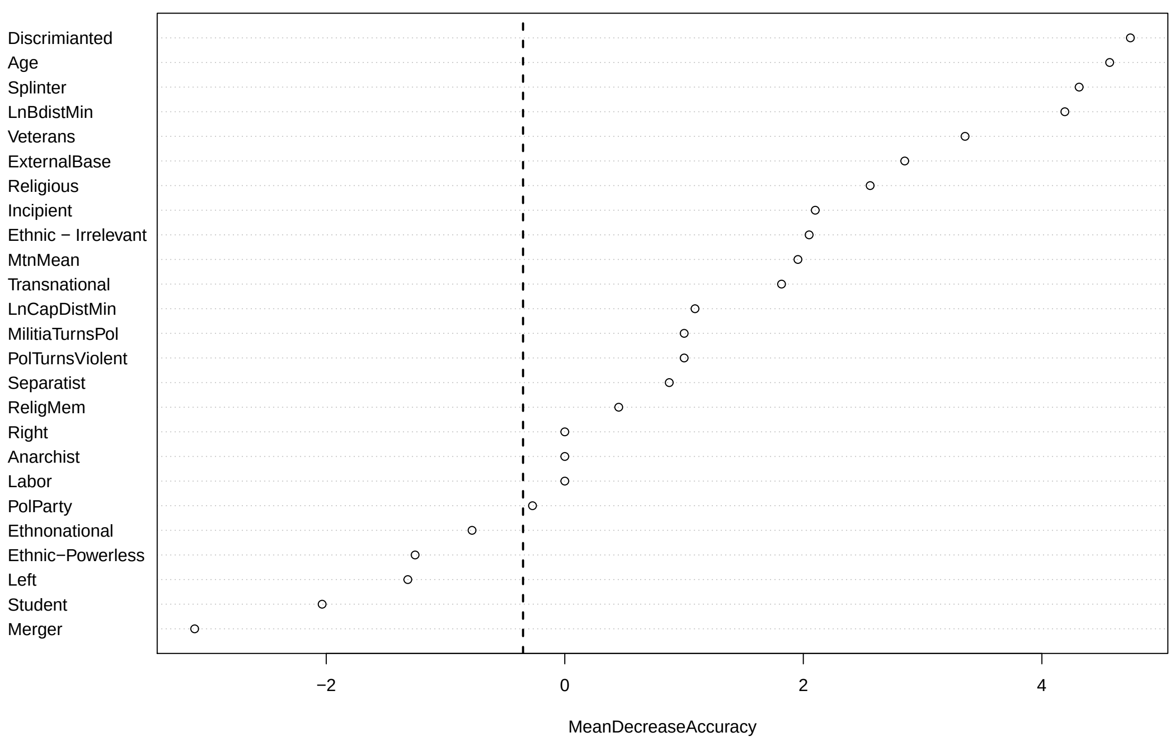


Figure 17: **Relative Variable Importance Plot for Strong States.** The dotted line is the relative importance of a random uniform variable added to the model as a placebo check. If a variable does better than this placebo, it is considered significant and falls to the right of the line. If a variable does worse than this placebo, it is considered insignificant and falls to the left of the line.



For comparison to Figure 17, I also include a relative variable importance plot for the weak states sample.

Figure 18: **Relative Variable Importance Plot for Weak States.** The dotted line is the relative importance of a random uniform variable added to the model as a placebo check. If a variable does better than this placebo, it is considered significant and falls to the right of the line. If a variable does worse than this placebo, it is considered insignificant and falls to the left of the line.



Hypothesis 3

When I rerun the hypotheses on a higher threshold, i find similar marginal rates of improvement on the basis of the out-of-bag error. Notably, the misclassification rate is much smaller because the higher threshold for civil war is much rarer.

Table 6: Summary of Marginal Prediction Improvement for Higher Civil War Threshold

Subset	Weak	Strong
No Information	0.897	0.945
RF	0.997	0.994
%Δ	0.100	0.049
Train N	170	551
Test N	85	212

Figure 19: Histogram of Marginal Improvement in Predictive Accuracy

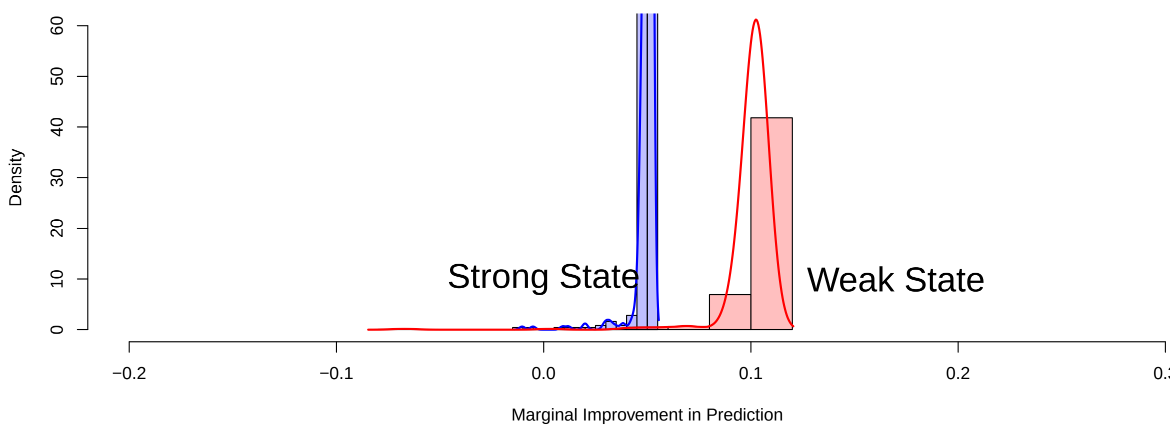


Table 7: **Cross-Sectional Data for Higher Civil War Threshold.** I use a linear probability model and choose the top five variables identified in random forest for each subset for inclusion. The dependent variable is whether an armed group causes a civil war that results in at least 1000-battle deaths.

	Model 1	Model 2	Model 3	Model 4
	Weak	Weak	Strong	Strong
DISCRIMINATED	0.32^{***}	0.36^{***}		
	(0.09)	(0.09)		
SPLINTER	0.07	0.14		
	(0.08)	(0.10)		
VETERANS	0.09	0.03		
	(0.06)	(0.08)		
LNAGE	-0.05^{**}	-0.06^{**}	-0.03^{***}	-0.03^{***}
	(0.02)	(0.03)	(0.01)	(0.01)
LNBDISTMIN	-0.03^{**}	-0.04	-0.01^{*}	-0.02^{*}
	(0.01)	(0.02)	(0.01)	(0.01)
CAPDISTMIN			-0.01	-0.01
			(0.01)	(0.01)
MTNMEAN			0.08^{**}	0.13^{***}
			(0.03)	(0.03)
LNINCIPIENT			0.02^{*}	0.02
			(0.01)	(0.01)
Country FE	N	Y	N	Y
R^2	0.23	0.36	0.04	0.12
Adj. R^2	0.20	0.27	0.03	0.05
Num. obs.	143	143	545	545
RMSE	0.29	0.28	0.23	0.23

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

I assess the magnitude and direction of these relationships using a linear probability model. Interestingly, the results suggest that visibility characteristics like age and border distance matter both in strong and weak states. Characteristics associated with strength like an armed group with veterans and combat experience just barely misses significance in the pooled model.

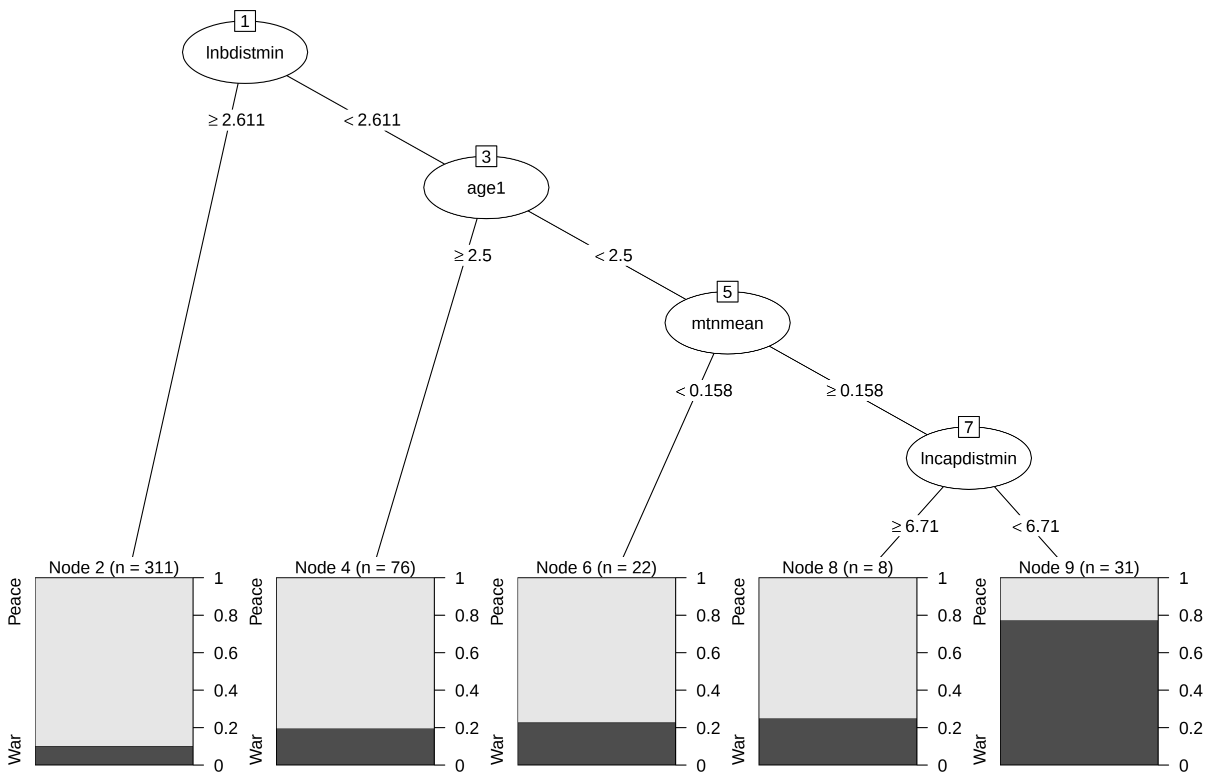
Alternate Specification: Weak versus Strong State Sample

The results in Section 5 might not be robust if GDP measures are endogenous to the formation of certain types of armed groups. In order to assess this, I run the results using an alternate sample of strong and weak states based on the UN Least Developed country criteria. Running these alternate results finds no difference.

Hypothesis 1 and 2

The results of a 6-fold CV Classification and Regression Tree for strong states is below.

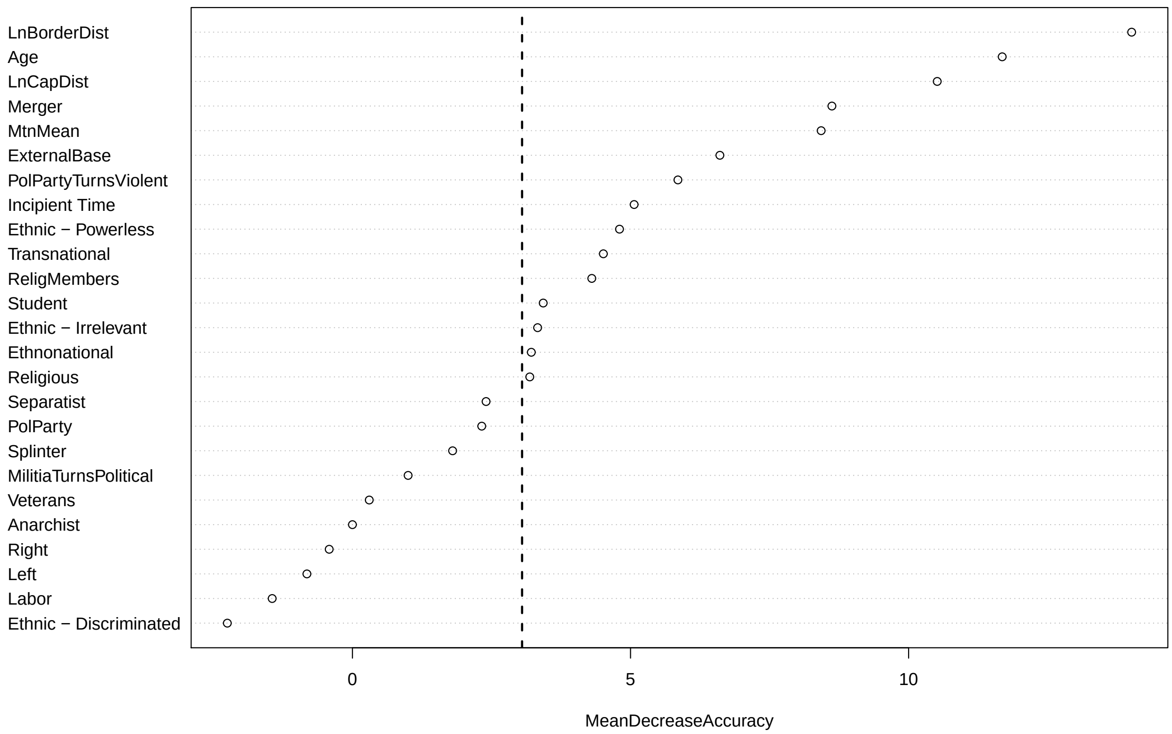
Figure 20: 6-fold CV Classification and Regression Tree for Strong States. The alternate state type specification shows a similar pruned CART as the original sample.



The results simplify to suggest very few organizational characteristics predict civil war. Further the same characteristics identified in the other sample emerge again.

The results are similar to the variable importance plot for the UN LDC. This sample down weights ideology more than the GDP sample, but the top five indicators are otherwise nearly identical.

Figure 21: **Relative Variable Importance Plot for Strong States.** The dotted line is the relative importance of a random uniform variable added to the model as a placebo check. If a variable does better than this placebo, it is considered significant and falls to the right of the line. If a variable does worse than this placebo, it is considered insignificant and falls to the left of the line.



Hypothesis 3

I compare the different prediction accuracy rates between the weak and strong state samples in Table 8. The results show that the marginal improvement in prediction is much higher for weak states (9.2%) than for strong states (1.1%). This again supports the prediction for Hypothesis 3.

I assess the magnitude of these relationships in a linear probability model. I take the top five relatively important variables from the random forest algorithm for the weak and strong subsets and run an OLS on the results. While age is statistically significant in all specifications for the GDP sample, it is only significant in the weak state even though the coefficient sign is in the right direction. The results on DISCRIMINATED, MERGER, LNBDISTMIN, and MTNMEAN are otherwise the same.

Table 8: Summary of Marginal Prediction Improvement for a UN LDC Subset

Subset	Weak	Strong
No Information	0.513	0.871
RF	0.605	0.882
%Δ	0.092	0.011
Train N	170	551
Test N	85	212

Table 9: **Cross-Sectional Data for Alternate Subset of Weak and Strong States.** I use a linear probability model and choose the top five variables identified in random forest for each subset for inclusion. The dependent variable is whether an armed group causes a civil war that results in at least 1000-battle deaths.

Subset	Weak Model 1	Weak Model 2	Strong Model 3	Strong Model 4
DISCRIMINATED	0.51^{***}	0.47^{***}		
	(0.09)	(0.04)		
LNAGE	-0.13^{***}	-0.12^{***}	-0.02	-0.02
	(0.03)	(0.04)	(0.01)	(0.01)
MERGER	0.29^{***}	0.28^{***}	0.14	0.15
	(0.09)	(0.09)	(0.14)	(0.14)
LNCAPDISTMIN	0.03	0.01	0.01	0.01
	(0.03)	(0.03)	(0.01)	(0.01)
MTNMEAN	-0.30^{***}	-0.21	0.12	0.22^{***}
	(0.11)	(0.13)	(0.08)	(0.08)
LNBDISTMIN			-0.05^{***}	-0.05^{***}
			(0.01)	(0.01)
Country FE	N	Y	N	Y
R^2	0.34	0.48	0.07	0.23
Adj. R^2	0.32	0.41	0.06	0.17
Num. obs.	165	165	523	523
RMSE	0.41	0.38	0.37	0.35

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

8.4 Appendix E: Sample Armed Group Narratives

Information about the armed group is coded from a narrative compiled by the author and a team of research assistants. I include two sample narratives from the Nigeria and Bangladesh cases, although the structure and length is comparable in every other cases. When information is ambiguous or not stated, it is coded as unclear or unknown.

Sample 1. Nigeria

Group Name: ODUA PEOPLES' CONGRESS

TORG ID: 951
Min. Group Date: 1995
Max. Group Date: 2011
Onset: NA

Aliases: Odua Peoples' Congress (Opc), Oodua Peoples Congress, Odua Peoples' Congress, Odua Peoples Congress, Oodua Peoples Congress, Yoruba O'odua Peoples' Congress, Yoruba Oodua Peoples Congress (Opc)

Part 1. Bibliography

- "The O'odua People's Congress (OPC): Fighting Violence with Violence." 2003. Human Rights Watch. https://www.hrw.org/report/2003/02/28/oodua-peoples-congress-opc/fighting-violence-violence

- Canada: Immigration and Refugee Board of Canada, Nigeria: Oodua People's Congress (OPC); leadership, membership, activities, and treatment by authorities (January 2005 - February 2006), 16 February 2006, NGA101048.E, available at: http://www.refworld.org/docid/45f1478a2f.html

- Canada: Immigration and Refugee Board of Canada, Nigeria: Information on the O'odua People's Congress (OPC), including the methods of recruitment and the selection of members; implications of being a member and the consequences of refusing to join (2003-May 2005), 9 June 2005, NGA100078.E, available at: http://www.refworld.org/docid/440ed7342.html

- "Oodua People's Congress." Para-Military Groups Africa. Global Security. N.D. http://www.globalsecurity.org/military/world/para/yoruba.htm

- Immigration and Refugee Board of Canada: The Oodua Peoples Congress (OPC), including when formed, leaders, aims, ethnicity, treatment of members and when it first came into conflict with the government authorities [NGA32576.E], 26. August 1999 (verfugbar auf ecoi.net) http://www.ecoi.net/local_link/188556/306595_de.html

- Yusuf Abdulazeez. 2013. O'odua People's Congress and the Changes in Nigeria's Political and Security Structures, Social Movement Studies, 12:2, 235-243,

- Yvan Guichaoua. 2006. "The Making of an Ethnic Militia:The Oodua People's Congress in Nigeria." Working Paper No. 26. Center for Research on Inequality, Human Security, and Ethnicity.
- Ed. Nicolas Florquin and Eric Berman. "Federated Niger Delta Ijaw Communities." Armed and Aimless: Armed Groups, Guns, and Human Security in the ECOWAS Region. 2005. Small Arms Survey. p. 330. `http://www.smallarmssurvey.org/fileadmin/docs/D-Book-series/book-01-Armed-and-Aimless/SAS-Armed-Aimless-1-Full-ma pdf`
- Canada: Immigration and Refugee Board of Canada, Nigeria: Location of branches of the Oodua Peoples Congress (OPC) other than in the states of Oyo, Ogun, Ondo, Osun, Kwara, Lagos, and Kogi; reports of violence involving the OPC in non-Yoruba areas, including the states of Abia and Cross Rivers (January - August 2001) , 24 August 2001, NGA37708.E , available at: `http://www.refworld.org/docid/3df4be828.html`
- GTD Perpetrator 20342. Global Terrorism Database. Study for Terrorism and Responses to Terrorism. Last Modified June 2016. `https://www.start.umd.edu/gtd/search/Results.aspx?perpetrator=20342`

Part 2. Basic Coding Changes
Aliases: Oodua Liberation Movement (OLM); Revolutionary Council of Nigeria
Group Formation: 1994 (form), unknown (first attack)
Group End: 2014 (last attack), 2017 (active as political movement)

Part 3. Narrative

Group Formation

The OPC formed in August 1994 in response to the government's cancellation of a 1993 Nigerian election where a Yoruba candidate appeared the likely winner (Human Rights Watch 2003; Florquin and Berman 2005, 330). The organization has an ethno-nationalist ideology, and claims that their motive is to safeguard the Yoruba culture and language. The OPC seeks autonomy for the Yoruba people, though it is not clear whether their political objective is the formation of a separate state, autonomy under the existing Nigerian government, or territorial reforms (Human Rights Watch 2003). The OPC's first attack is not officially recorded. By 1999, the group had already conducted several attacks against other ethnic groups and clashed with the police repeatedly (Human Rights Watch 2003).

Geography

The OPC has known bases in southwest Nigeria which has a tropical environment: Oyo, Ogun, Ondo, Osun, Kwara, Lagos, and Kogi (Canada IRB 2001). A majority of attacks and clashes happen with the police in Yoruba areas like Ikorodu and Lagos (GTD 2016). Some records suggest that the OPC was also active in Llorin, the capital of Kwara, around

October and November 2000 (Canada IRB 2001). The OPC is not a transnational group.

Organizational Structure

The OPC had several leaders including Dr. Frederick Fasehun, a doctor with no prior political experience (Human Rights Watch 2003). In 1999, the OPC formed two distinct wings, the moderate wing, which was led by Fasehun and was willing to engage in politics. The more radical wing was led by Gani Adams, a previous carpenter, and frequently used violence (Human Rights Watch 2003; Florquin and Berman 2005, 330). The leaders of the OPC tended to have a higher level of education and political experience, while the members have minimum education and are typically young, unemployed men (Human Rights Watch 2003; Florquin and Berman 2005, 330). A majority of OPC members are Yoruba. While it is not clear if the OPC collects fees from regular members, Yoruba politicians that use their OPC membership as a platform to gain support have to pay the group (Florquin and Berman 2005, 330). The group is organized into a hierarchy composed of branches, zones, and wings. The Annual National Conference is at the forefront of all major group decisions, and the National Executive Council is its governing body (Human Rights Watch 2003). There are various estimates of the OPC's size, but most agree that each of the 20 battalions or zone had approximately 200 men (Florquin and Berman 2005, 330).

External Ties

Many local governments in Yoruba majority areas hired the OPC as a vigilante justice organization (Canada IRB 2006). In addition, Yoruba politicians may have partially financed the OPC's activities (Florquin and Berman 2005, 330; Canada IRB 2006). The governor of Lagos allegedly hired the OPC as a para-military force to help his 2003 victory (Florquin and Berman 2005 p.330). In 1999, the OPC split into two opposing wings, a moderate wing led by its founder Fasehun, and a radical militant wing, which was led by Gani Adams (also known as the Oodua Liberation Movement, or the Revolutionary Council of Nigeria) (Florquin and Berman 2005 p.330; Global Security N.D).

Group Outcome

A government anti-crime force called "Operation Sweep" in December 1998 in Bariga, Lagos resulted in the death of four members of the OPC (Global Security N.D). In 1999, the Nigerian government banned the OPC, and ordered police officers to shoot on sight. This resulted in a series of violent confrontations in Yoruba territory (Canada IRB 2006). In November 2005, Nigerian police arrested both of the group's leaders (ibid.). After a spell of non-activity between 2005 and 2013, the organization began using violence again in 2013. Its last suspected attack was in 2014 when it attacked the Nigeria Prisons Service in Ado-Ekiti (GTD 2017). The group is still active in Yoruba politics as of 2017.

Sample 2. Bangladesh

Group Name: CHITTAGONG HILL TRACTS UNITED PEOPLE'S PARTY

TORG ID: 1441
Min. Group Date: 1972
Max. Group Date: 2009
Onset: 1975

Aliases: Parbatya Chattagram Jana Sanghati Samity (Pcjss) - Bangladesh, Chittagong Hill Tracts United Peoples Party, Chittagong Hill Tracts United People's Party, Parbatya Chattagram Jana Sanghati Samiti, Parbatya Chattagram Jana Sanghati Samity, Parbatya Chattagram Jana Sanghati Samity (Pcjss)

Part 1. Bibliography

- Kristian Gleditsch, David Cunningham, and Idean Salehyan. "Bangladesh vs JSS/SB." Non-State Actor Data. 2013. `http://privatewww.essex.ac.uk/~ksg/data/NSAEX_casedesc.pdf`

- Aftab Ahmed. "Ethnicity and Insurgency in Chittagong Hills Tract Region." Journal of Commonwealth and Comparative Politics, 2008, `http://www.tandfonline.com/doi/abs/10.1080/14662049308447665?journalCode=fccp19`

- Amnesty International. "Bangladesh: Unlawful Killings and Torture in the Chittagong Hill Tracts." 1986. `https://www.amnesty.org/en/documents/asa13/021/1986/en/`

- "Chittagong Hill Tracts Political Parties." Global Security n.d., `http://www.globalsecurity.org/military/world/bangladesh/political-parties-cht.htm`

- "Bangladesh: A Critical Review of the CHT Peace Accord." UNDP Working Paper No. 2. Regional Centre in Bangkok. 2005. `https://www.academia.edu/2504753/Bangladesh_A_Critical_Review_of_the_Chittagong_Hill_Tract_CHT_Peace_Accord`

- "Report on the Status of Implementation of the CHT Accord." Chittagong Hills Tract Commission. 2013. `http://www.chtcommission.org/Report-on-Impln-of-CHT-Accord-January-` `pdf`

- "Chittagong Hills Tribe in Bangladesh," Minorities at Risk. Last Updated 2006. `http://www.mar.umd.edu/assessment.asp?groupId=77101`

- Kazi Montu. "Tribal Insurgency in Chittagong Hills Tract." Economic and Political Weekly. 1980. JSTOR.

- Patrick Brogan. *World Conflicts: A Comprehensive Guide to World Strife since 1945* Scarecrow Press. 1999

- Mohammad Fakhrus Salam and Hajera Aktar. "Ethnic Problems in Bangladesh: A Study of Chittagong Hill Tracts." SUST Journal of Social Sciences. 2014. `http://sustjournals.org/uploads/archive/aba5f3414b6bb5d9145e7a8762c8d6a2.pdf`

- Zarein Ahmed. "Shanti Bahini." IPCS. 1999. http://www.ipcs.org/article/bangladesh/shanti-bahini-293.html
- Bushra Chowdhury. "Issues of the Implementation of the Chittagong Hills Tract." 2002. https://www.ideals.illinois.edu/bitstream/handle/2142/228/chowdhuryop.pdf?sequence=2
- Mark Levene. "The Chittagong Hill Tracts: A case study in the political economy of 'creeping' genocide." Third World Quarterly. 1999. 20(2): 339-369.
- Caf Dowlah, "Jumma Insurgency in Chittagong Hills Tract: how serious is the threat to Bangladesh's national integration and what can be done?" Small Wars and Insurgencies, Vol 24(5). 2013.
- GTD Perpetrator 595. Global Terrorism Database. Study for Terrorism and Responses to Terrorism. Last modified June 2017. http://www.start.umd.edu/gtd/search/Results.aspx?perpetrator=595

Part 2. Basic Coding Changes
Additional Aliases: Shanti Bahini
Group Formation: 1972 (form), 1975 (Attack)
Group End: 1997 (disarm)

Part 3. Narrative

Group Formation

PCJSS was formed in 1972 by members of the Chittagong Hills Tract (CHT) tribe in eastern Bangladesh (UNDP 2005; Salam and Aktar 20154, 55). They created an armed wing, the Shanti Bahini, in 1973 (UNDP 2005). They were an ethno-nationalist separatist groups that claimed the Chittagong Hills District after settlers began to immigrate to the area from other parts of Bangladhes. The Prime Minister of Bangladesh, Mujib, rejected the claims and threatened to send even more settlers into the region (UNDP 2005). Initially, the group was very hopeful it could bargain a negotiated settlement with the government, but eventually started their armed campaign after Mujib was assassinated in 1975 (UNDP 2005). The group partially ascribed to Marxist-styles of warfare as seen by the type of guerrilla tactics employed, although they were not a leftist organization (Ahmed 1999).

Geography

The group operated throughout the Chittagong Hill District, but primarily operated in three districts: Rangamati, Bandarban, and Khagrachari, all located in the CHT district (Chowdhury 2002; Ahmed 1999). There is no evidence it ever conducted any transnational attacks (GTD 2017). Manobendra Narayan Larma fled to India after Mujib's assassination in 1975 and organized the insurgency from an external base there, but it is unclear whether this move predates or postdates the start of the insurgency (UNDP 2005; Salam and Aktar 2014, 57).

Organizational Structure

PCJSS leader was Manobendra Narayan Larma who fled to India after Mujib's assassination and organized the insurgency from India (UNDP 2005; Salam and Aktar 2014). His brother Manobendra Narayan also helped him lead the group (Ahmed 1999). The group was composed of ethnic CHT (Gleditsch et al. 2013, 434). MAR argues the group has a "high level of group organization and cohesion," but provides no evidence why (MAR 2006). The group's armed wing was divided along several different fields including a medical wing, communication, and tech teams (Ahmed 1999). It studied Marxist guerrilla tactics and appointed J.B. Larma the leader (Ahmed 1999) UCDP estimates the group had 2000-7000 guerrillas "during the course of the conflict" (Gleditsch et al. 2013, 434). The group primarily funded itself through extortion and taxation of officials (Ahmed 1999).

External Ties

In late 1975 or early 1976, India began providing external support to the PCJSS and SB under the Research and Anaysis Wing (RAW) (Ahmed 1999; Gleditsch et al. 2013, 434). SB had an external base in Tripura, received military training from Indian security forces, and also had Indian weaponry (UNDP 2005). India later denied these claims (Gleditsch et al. 2013, 434).

Group Outcome

As early as 1973, the government began increasing its military response in the area. The JSS formed in response to "came in the wake both of Sheik Mujib's refusal to consider autonomy and of a series of sweeping and indiscriminate reprisal raids against the CHT" (Leven 1999, 357). Bangladesh banned the PCJSS political group and led to the creation of the armed wing (Leven 1999; Dowlah 2013). The government responded to the group in 1975 with a massive military campaign including a full-scale occupation of the CHT (UNDP 2005). These state-initiated actions match the UCDP's start date of the conflict. When Rahman came to power in 1976, he deployed 150,000 military personnel to the region, which forced the JSS/SB to go underground and increase their level of effort against Rahman (Dowlah 2013). Larma was arrested in 1975 (Ahmed 1999). It also tried to make some accommodations including the creation of the Chittagong Hills Development Board in January 1976 (Chowdhury 2002). A conflict in the PCJSS leadership led to the assassination of Larma on November 10, 1983 (Chowdhury 2002).

Talks between Zia and the PCJSS fell through repeatedly during his tenure due to a lack of trust between parties (UNDP 2005). The government continued its militant, hard-line approach until 1991 when the BNP - a civilian government - initiated peace talks and began to treat the insurgency as a political movement (UNDP 2005). An eventual peace accord was reached on Dec. 2, 1997 granting the CHT some greater autonomy and a fixed number of seats in the Parliament (UNDP 2005). The group's last violent attack was in 1997 (GTD 2017).