Not Just Another Armed Actor: Explaining Pro-Government Militia Mobilization and Participation in Civil Conflict

by

Devin P. Brown

A Dissertation Submitted to the
Graduate School of Social Sciences and Humanities
in Partial Fulfillment of the Requirements for
the Degree of

Master of Arts

in

International Relations

KOÇ ÜNİVERSİTESİ

June 2021

Not Just Another Armed Actor: Explaining Pro-Government Militia Mobilization and Participation in Civil Conflict

Koç University
Graduate School of Social Sciences and Humanities
This is to certify that I have examined this copy of a master's thesis by

Devin P. Brown

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Assoc. Prof. Belgin San-Akca (Advisor)

Asst. Prof. David Carlson [Co-Advisor]

Dr. Bülent Aras

Date:

To Istanbul.

ABSTRACT

Not Just Another Armed Actor: Explaining Pro-Government Militia Mobilization and Participation in Civil Conflict
Devin P. Brown
Master of Arts in International Relations
June 2021

Pro-government militias (PGMs) are a critical yet undertheorized actor in civil conflicts. Current research has not considered what factors lead to PGM participating in internal conflict, leaving a gap in our understanding of the ultimate purpose of PGMs. I test this relationship through the lens of two competing hypotheses: PGMs as force multipliers and PGMs as intelligence gatherers. Using data from the Uppsala Conflict Data Program and the Pro-Government Militias Database, I use an Andersen-Gill hazard model and negative binomial regression as well as out-of-sample predictive tests to determine when states will mobilize PGMs and what leads states to utilize multiple PGMs. Results show that the need for intelligence drives states to mobilize new PGMs while both latent military threat and the need for local information contribute to multiple PGM usage. I argue that these results require a revised approach to studies of PGMs, focusing on their intelligence gathering purpose rather than treating them as "just another armed actor" in civil conflict. This revision has implications not only for studies of pro-government militias but also for larger, integrated studies of civil conflict.

ÖZETÇE

Sadece Başka Bir Silahlı Aktör Değil: Hükümet Yanlısı Milislerin İç Çatışmalarda Mobilizasyonu ve Katılımını Açıklamak

Devin P. Brown
Uluslar Arasıİ lişkileri, Yüksek Lisans
Haziran 2021

Hükümet yanlısı milisler (HYM) iç çatışmalar bakımından önemli oyuncular olsalar da haklarında yeterince teori geliştirilmemiştir. Mevcut araştırmalar HYMlerin iç çatışmalara katılmasına yol açan unsurları incelememekte, bu da HYMlerin nihai amacına dair anlayışımızda bir boşluğa sebep olmaktadır. Bu ilişkiyi iki farklı hipotezin objektifinden test etmekteyim: kuvvet çarpanı olarak HYMler ve istihbarat toplayan HYMler. Uppsala Çatışma Veri Programı ve Hükümet Yanlısı Milisler Veri Setinden veri alarak, HYMleri hangi devletlerin harekete geçirdiğini ve devletlerin birden fazla HYM kullanmasına sebep olan unsuları belirlemek için örnlekem dışı tahmin testlerinin yanı sıra Andersen-Gill risk modelini ve negatif binom regresyonunu kullanmaktayım. Sonuçlar, istihbarat ihtiyacının devletleri HYM kullanmaya ittiğini, bununla birlikte, gizli askeri tehditlerin ve yerel bilgi ihtiyacının birden fazla HYM'nin kullanılmasına sebep olduğunu göstermektedir. Bu sonuçların, HYMlere dair kuramda, HYMleri iç çatışmalarda "sadece başka bir silahlı aktör" olarak değerlendirmektense, HYMlerin isthbarat toplama amacına odaklanan bir yaklaşım değişikliği gerektirdiğini tartışmaktayım. Bu yalnızca HYMlere ilişkin akademik çalışmalar bakımından sonuç doğurmayacak ve iç çatışma konulu daha kapsamlı ve entegre akademik çalışmalara da etki edecektir.

ACKNOWLEDGMENTS

I would like to thank numerous people for their help throughout this project. To David Carlson for his unwavering faith in me. To Belgin San-Akca for challenging me to think carefully about theory and allowing me the freedom to follow my own ideas. To Yeşim for listening, for encouraging me, and for the hugs. To Merve and Kağan for Friday-night Okey and Settlers. Lastly, to my Dad and Mom for their constant love and unrelenting support, even from an ocean away.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

PGM	Pro-Government Militia
NAG	Non-State Armed Group
UCDP	Upsala Conflict Data Program
PGMD	Pro-Government Militias Database
NSA	Non-State Actors Dataset
PC/PCA	Principal Component (Analysis)
PKK	Kurdistan Worker's Party
RUF	Revolutionary United Front
IS	Islamic State
AFMC	Armed Forces Military Council
LTTE	Liberation Tigers of Tamil Eelam

Chapter 1

INTRODUCTION AND THEORETICAL GROUNDING

In the past decade, conflict scholars have come to understand that civil conflicts are rarely dyadic affairs. A body of literature has considered the role of external state support for insurgent groups (Gleditsch 2007; Salehyan, Gleditsch, and Cunningham 2011; San-Akca 2016), while another group of studies have examined rebel group fractionalization and competition (Akcinaroglu 2012; Gade et al. 2019). As the conception of a unitary rebel side expanded, researchers began theorizing about the state side, specifically regarding the use of paramilitaries. Scholars have found that pro-government militias (PGMs) are ubiquitous participants in civil war. 43% of PGMs listed in the Pro-Government Militias Database are present during some type of civil conflict (Carey, Mitchell, and Lowe 2012).

Although the existing literature agrees that most PGMs serve purposes outside of conflict, this project will focus more narrowly on PGMs as conflict actors. Extant work on PGMs focuses on militias' behavior and on PGM's broader effects on conflict duration and conflict dynamics. What little research has been done on PGM formation is primarily qualitative. The only quantitative treatment of paramilitary mobilization covers all PGMs both in conflict and non-conflict settings and uses regime type as the central explanatory variable (Carey, Colaresi, and Mitchell 2015). This leaves a large gap in the literature. What drives PGM participation in civil war?

I divide this question into two parts. First, I examine the factors associated with initial PGM mobilization, and second, I analyze what leads states to use multiple PGMs. I approach these puzzles through two competing explanations. One hypothesis argues that stronger rebel groups and greater latent military threat drive the state to mobilize PGMs as force multipliers. The other hypothesis posits that when states have difficulty gathering reliable information about insurgent groups, the state will utilize PGMs as intelligence gatherers. I test both hypotheses quantitatively using a specialized hazard model to estimate PGM mobilization and a negative binomial regression to model multiple PGM participation in conflict as well as out-of-sample predictive tests. The results suggest that the need for intelligence primarily drives PGM mobilization, while a combination of both hypotheses explains multiple paramilitary involvement in civil conflict.

This thesis contributes to ongoing research in multiple ways. It is the first quantitative analysis that uses conflict-specific factors to explain PGM formation and participation in civil conflict. Furthermore, the results suggest that an overemphasis on PGMs as violent actors may misrepresent their actual purpose in conflict and lead to misguided conclusions and theoretical expectations. Rather than treating PGMs as "just another armed actor," research should recognize the centrality of local intelligence in a state's decision to utilize PGMs in internal conflict. This project also illustrates the theoretical usefulness of flexible modeling strategies and predictive tests.

Finally, this analysis highlights some of the theoretical and methodological deficiencies that will hinder future research. First, "pro-government militia" is too broad of a category to allow for specific and incisive theory. The variation between groups is so great that it prevents meaningful analyses and theory-building. While international relations as a discipline has made great methodological strides in the past few years, there are still many opportunities to improve existing methods to better reflect complex causal processes.

The project proceeds as follows. The rest of this chapter begins with a discussion on defining pro-government militias and their theoretical importance in political science. It continues with a survey of the current scholarship on PGM behavior and formation. Chapter 2 outlines the hypotheses, the data sources, dataset adjustments, and the modeling strategy. Then it reports the results from the preliminary hypothesis tests. Chapter 3 further explores the two hypotheses. First, it tests whether a slightly different specification of the military threat variable would yield better results than Chapter 2's conceptualization. Then it explores the impact of two strong, yet theoretically ambiguous variables to determine whether they support one hypothesis more than the other. Lastly, it runs regressions with all possible combination of variables and tests the predictive accuracy of each combination. This final test determines whether both hypotheses retain explanatory power over the outcome variable or if one hypothesis does a better job. Chapter 4 outlines in detail the theoretical contributions of this analysis, areas for improvement, and opportunities for future research.

1.1 Defining a Pro-Government Militia

Defining a pro-government militia is a difficult task. Many scholars have written broadly about "paramilitaries" (Jentzsch, Kalyvas, and Schubiger 2015), but many distinguish between "death squads," (Campbell and Brenner 2002) "self-defense forces," (Mazzei 2009), "state-sponsored militias," (Ahram 2011b) and "civil defense forces" (Clayton and Thompson 2016). Each of these characterizations refers to a narrower type of pro-government militia with their attendant purposes, illustrating the problems inherent in any large-N analysis of these groups. A few classification schemes disaggregate these groups according to their ideology (Staniland 2015), conflict experience (Biberman 2018), and connection to the state (Carey, Mitchell, and Lowe 2012). Within large, cross-group comparisons however, Carey, Mitchell, and Lowe's (2012) definition has become the standard. The authors outline four criteria for inclusion as a PGM in their dataset:

1. The group must be identified as supporting or sponsored by the government.
2. It must not be part of the regular armed forces.
3. It must be armed.
4. It must have some form of independent organization.

This definition exhibits clear similarities to how scholarship defines rebel and insurgent groups, but the key distinction between a PGM and a non-state armed group (NAG)[1] lies its relationship to the state in which it operates. PGMs have the approval of the domestic government. In many cases the state supports and cooperates with the militia, but explicit material support isn't strictly necessary for a group to be coded a PGM. A few short examples will illustrate this definition.

The Village Guard in Turkey is one of the most common and clear examples of a pro-government militia. Early in the conflict with the Kurdistan Worker's Party (PKK[2]), the Turkish government created local paramilitary organizations to help combat the PKK and gather information on PKK movements and sympathetic civilians. The village guards regularly cooperated with the Turkish military, sharing intelligence and participating in joint operations. By the early 2000s, Turkey was paying 60,000 members of the village guard network (Belge 2016). This group was government-initiated, coordinated, and supported like many other PGMs. Around 2000, the village guards were formally integrated into the Turkish security apparatus, transforming it from a PGM into an official arm of the police.

A slightly more complicated example is the Peshmerga Kurdish Militia in Iraq. Originally formed after the collapse of the Ottoman Empire, the Peshmerga are the military wing of Kurdish political organization. From the 1970s to the turn of the century, the Peshmerga fought against the Iraqi government in a bid for independent statehood. After the US invasion in 2002, the Kurds gained regional autonomy, and the Iraqi government began allowing the Peshmerga to handle security in Kurdish areas, transforming this erstwhile NAG into a PGM. Supported by the United States, the Peshmerga was critically important in the fight against the Islamic State (IS) and was the primary fighting force responsible for recapturing the erstwhile IS capital, Mosul, in 2017 (BBC 2014). In contrast to the Village Guards, the Peshmerga formed independently and currently receives material support from a foreign government. The critical attribute that designates this group as a pro-government militia is the permission of the Iraqi government to operate in Iraq and the militia's connection to the state through the Kurdistan Regional Government.

---

[1] This terminology was originally used in San-Akca (2009)

[2] Partîya Karkerên Kurdistanê in Kurdish.

Finally, the Revolutionary United Front (RUF) in Sierra Leone is briefly coded as a PGM. Similar to the Peshmerga in Iraq, the RUF fought against the Sierra Leone government in a 9-year civil war starting in 1991. In 1997, the Armed Forces Military Council (AFMC) executed a coup d'état and invited the leader of the RUF to join the new government (Gberie 1997), transforming the RUF into a pro-government militia. After the AFMC was removed one year later by an international coalition, the RUF reverted back to its insurgency. During the short tenure of the AFMC, the RUF operated with the permission and support of the government and therefore, qualifies as a PGM, even if only for a year.

These examples highlight the different forms that a PGM can take. In some cases, an insurgent group can be utilized as a pro-government militia. In each instance, however, a PGM operates with the consent of the domestic government. In this paper, I will interchangeably refer to these groups as PGMs, militias, or paramilitaries.[3]

1.2 The Theoretical Importance of PGMs

Pro-government militias are theoretically important for numerous reasons. First, as will be discussed in the following section, paramilitaries commonly operate in civil conflicts. Since studies of civil wars began, the understanding of internal conflicts as dyadic and domestic affairs has slowly eroded. The literature has expanded to include external state sponsors of non-state armed groups (Gleditsch 2007; Gleditsch, Salehyan, and Schultz 2008; Salehyan, Gleditsch, and Cunningham 2011; San-Akca 2016), IGOs like the United Nations (Tierney 2005), and multiple insurgent groups, including alliances between different NAGs (Akcinaroglu 2012; Gade et al. 2019). Most studies, however, consistently treat the state as the sole protector of its interests. Paramilitary studies have shown that this is not the case. Many states rely on militias and other irregular forces during internal conflict.

[3]Most literature on PGMs also uses these terms interchangeably, but 'paramilitary' has also been used as a term for irregular or military police forces. For a discussion of this distinction see Böhmelt and Clayton's (2018) article, "Auxiliary Force Structure."

This fact raises questions about a state's ability to negotiate peace agreements, the potential accountability of the government for the actions of PGMs, and ultimately about sovereignty and the monopoly on violence in a territory. First, introducing another party into a final peace agreement can generate significant negotiation problems, especially if that new party holds a veto over the final agreement (Cunningham 2006). One study found that paramilitaries operate as "no peace, no war proponents," favoring long-term, low-intensity conflicts to ensure the group's survival and access to resources (Aliyev 2019). For example, at the end of the Yugoslav Wars, Serbia's erstwhile paramilitary, the Arakan Tigers, lost access to human trafficking and arms trade networks and collapsed as a result. Even when PGMs cannot veto a peace agreement, violence and atrocities committed by paramilitaries during the conflict can exacerbate distrust of the government. If the state refuses to disarm PGMs after a conflict, paramilitaries can create commitment problems, obstructing a final peace agreement (Aliyev 2019). In sum, PGMs influence conflict dynamics as well as conflict resolution efforts, requiring special attention in numerous other literatures.

Second, a critical debate within paramilitary studies centers around the principal-agent relationship between the state and the PGM. Can or should states be held accountable for the actions of PGMs? Are PGMs simply an extension of state authority, or do they retain enough agency to eliminate the state's liability? Carey, Colaresi, and Mitchell (2015) argue that weak democratic states will use PGMs that are loosely connected to the government to intimidate opponents and to avoid international accountability for violence, while consolidated autocracies prefer a stronger connection to paramilitaries to better control PGM behaviors. This tension between distance and control is also present in the selection of PGMs. Staniland (2017) highlights how ideological affinity between the state and pro-government militia provides the state with more control over PGM behaviors. Alternatively, trained and seasoned militia groups may be more effective agents but are harder to control than less experienced groups (Biberman 2018). Ultimately, the literature must answer the question: Are states *unable* or are they *unwilling* to restrain paramilitary groups?

These questions highlight the final theoretical implication of paramilitaries: their impact on state sovereignty. Classical definitions of the state critically rely on the idea of the monopoly of legitimate violence in a territory (Tilly; Weber 1946). Empirical studies of paramilitary involvement in states reveals that, in fact, many countries exhibit an "oligopoly of violence" (Carey and Mitchell 2017). Does this require a revision of the common conception of statehood, or can we adjust existing definitions to account for this phenomenon?

Many popular revisions of sovereignty focus on post-colonial, especially African, states when discussing the failure to achieve the monopoly of violence in a territory (Clapham 1996; Jackson 1990). Scholars have even focused particular attention on how paramilitaries in fragile states exacerbate insecurity (Andersen, Moller, and Stepputat 2007). However, these discussions suffer from two particular shortcomings when applied to the question of pro-government militias. First, PGMs are not solely third world phenomena. Irregular forces such as paramilitaries or mercenary groups were commonly used during European state consolidation (Pereira 2003; Campbell and Brenner 2002). Tilly's war-making thesis oversimplifies the complex negotiations and interactions that resulted in modern monopolies on violence. Even during the Cold War, however, multiple European states allowed a NATO "stay behind army" named Gladio to operate in their territory, rooting out communist sympathizers and carrying out targeted killings, kidnappings, and bombings (Ganser 2005). Even though these tactics are more prevalent in low-capacity or weak states, they are not exclusive features of new or underdeveloped states

This emphasis on the monopoly of violence also treats it as an achieved and permanent condition. States must continually assert and protect their monopoly (Campbell 2002; Pereira 2003). Sovereignty is a dynamic process involving negotiation with multiple levels of society and power holders. At any point, a state may subcontract certain sovereign functions to private actors.[4]. This has been a hallmark modern statehood as countries struggle to keep pace with technological advancements and decreasing revenues (Kaldor 2012). Especially for newer states,

[4]This is not limited to security. States also use private actors to provide services or goods to their population. See Donahue and Zeckhauser (2011).

its obligations may exceed its capabilities, requiring the state to delegate sovereignty in order to function. A segment of literature has examined how by tying foreign investment to state performance on human rights, international aid organizations and IGOs actually incentivize leaders to use PGMs to maintain access to these funds while avoiding accountability for abuses (Andersen, Moller, and Stepputat 2007; Ahram 2011a).

Other research has explored how PGMs operate as a critical provider of public goods, however (Biberman and Turnbull 2018). In his influential book *The Logic of Violence in Civil War* (2006), Kalyvas argues that militias are used to *enhance* state control over territories and to build state capacity. Specifically he provides an example from Guatemala in which local militias fulfilled judiciary functions during the civil war. The literature is divided on the actual effect of paramilitaries on state sovereignty, but clearly, classical conceptions sovereignty no longer reflect the reality of sovereign governance. The monopoly on violence has been replaced by public private partnerships (Davis 2009) and pro-government militias are a critical yet undertheorized actor in these partnerships.

Although tangential to this project, these larger theoretical discussions illustrate the importance of research on paramilitaries, not only for studies of civil conflict but even for foundational debates in political science. The following section will address the current state of the literature specifically on PGMs.

1.3 The Ubiquity of PGMs and Current Debates

In recent years, PGMs have become a serious topic in the literature on civil conflict. The main dataset—the Pro-Government Militias Database (PGMD)—records 330 different PGMs from 1981–2007. Although militias are generally associated with conflict, the PGMD finds that of their coded PGMs, only 43% of groups are active during conflict years, while 57% are active in non-conflict years (Carey and Mitchell 2017), challenging this basic assumption. Most groups have an informal connection to the government; thus, these groups operate outside of the state's official chains of command. A key finding from this dataset was that both authoritarian and democratic, rich and relatively poor states utilize PGMs.

Existing studies of PGMs have centered around three themes: why and how governments cooperate with PGMs and how PGMs impact conflict dynamics. Regarding the purpose of PGMs, many governments rely on militias to help in counterinsurgency wars as intelligence gatherers. Having been recruited locally, PGM members have an intimate knowledge of local politics and can relay valuable information about insurgent activity and civilian collaborators to the government (Jentzsch, Kalyvas, and Schubiger 2015; Lyall and Wilson 2009). PGMs may also be used as force multipliers when a state is faced with both domestic and international threats. A within-case study of Myanmar showed that the combination of external and internal threats leads governments to use PGMs to suppress domestic rebellion and reserve conventional forces for external enemies (Eck 2015). Outside of conflict, PGMs can be used as "coup-proofers' (De Bruin 2018), operating in parallel with the conventional military but fiercely loyal to a particular leader or regime. The Revolutionary Guard in Iran is a clear example of this type of PGM.

The relationship between the government and a PGM has also been a popular topic in the literature. Most studies utilize the principal-agent framework to analyze how these actors interact. This approach became common in conflict studies with Salehyan's (2010) analysis of external state support for armed groups and has been fruitfully utilized throughout political science. A group of researchers have sought to answer how can governments delegate violent tasks to internal armed groups without the group become a security threat itself. Biberman argues that skilled "veteran" groups are significantly more effective than "rookie" groups, but more difficult to control. The government's job, thus, becomes managing the trade-off between effectiveness and control (Biberman 2018). Another scholar has found that PGMs which operate in state controlled areas are more likely to selectively use violence while those operating in rebel-controlled areas engage in more predatory behavior (Barter 2013). Stanton (2015) asserts that when PGMs are recruited from the same constituency as the insurgent group, they are less likely to target civilians.

States may also take advantage of this principal-agent problem to avoid being held responsible for violence against political opposition or civilians. Especially when weak democracies receive substantial support from democracies, governments will seek to avoid accountability for internal violence by delegating these tasks to PGMs. Authoritarian states, however, prefer more formal ties with internal militias, providing a greater level of control over PGM activities but also incurring responsibility for the militias' actions (Carey, Colaresi, and Mitchell 2015).

Other studies argue that the principal-agent problem is overstated, however. One analysis found that PGMs rarely engage in more severe violence than the state. When the state targets civilians, PGMs also target civilians but when the state avoids civilian victimization, so does the PGM (Stanton 2015). Relatedly, Cohen and Nordås (2015) found that PGMs engage in sexual violence more when the state trains them, refuting arguments that extreme violence follows the logic of delegation or indiscipline. These examples illustrates the ongoing debate on how states use and interact with paramilitaries.

Finally, studies on paramilitaries have considered the effects of PGM involvement in civil conflicts. One of the first quantitative analyses of PGMs found that an active PGM increases conflict incidence (Carey, Mitchell, and Lowe 2012). As mentioned previously, by delaying or obstructing a potential peace agreement, PGM presence extends conflicts (Jentzsch, Kalyvas, and Schubiger 2015; Abbs, Clayton, and Thomson 2020; Aliyev 2020). When a PGM is active in a conflict, the conflict is more likely to terminate in a long-term ceasefire or transition to continuous, low-intensity violence (Aliyev 2019). PGM involvement in civil conflict also leads to higher death tolls, both for combatants and for civilians (Clayton and Thomson 2014; Clayton and Thompson 2016; Abbs, Clayton, and Thomson 2020), except in situations where the PGM is recruited and operates locally. In these cases, social ties to the community restrict indiscriminate violence against civilians (Stanton 2015). Finally, when states initiate or support local civilian defense forces, they are more likely to decisively defeat an insurgent group (Peic 2014).

A group of scholars have also examined the intersection of ethnicity and paramilitary groups. PGMs that share an ethnicity with the targeted insurgent group are better able to identify rebels and issue credible threats to non-compliant civilians (Lyall 2010; Staniland 2012). When a regime and PGM share an ethnicity, the government is able to exert tighter control over PGM behavior, but this also tends to heighten ethnic tensions in a state, leading to more intense violence and longer conflicts (Abbs, Clayton, and Thomson 2020).

1.4 PGM Mobilization in the Literature

So far, this survey of the literature has focused on the behavior of PGMs, their relationship with the government, and their impact on conflict dynamics, but this still leaves one central question unanswered: why do these groups form in the first place? Existing studies on this question are mostly qualitative, focusing on a specific region or state and enumerating complex causal chains. These studies can be divided into two broad causes: structural factors and civilian agency.

1.4.1 Structural Explanations of PGM Mobilization

Structural explanations of PGM formation usually focus on situations that compel a government to create and utilize pro-government militias. Ahram's study focuses on former colonial territories and seeks to explain variation in a late-developing state's (LDS) choice of military strategy. He argues that two central variables determine whether a state will rely on PGMs for security: the type of decolonization (either revolutionary or through a bureaucratic transfer of power) and the external threat environment. Where states successfully rebel against their former colonizer and face few external military threats, the new, revolutionary government disposes of the former colonial military apparatus and relies on their erstwhile revolutionary connections to create PGMs that act as the primary coercive force in the country. When these LDS face external threats, however, the environment compels states to centralize authority and establish a standing army, similar to Tilly's (1992) expectations. Where there is a peaceful transfer of power, the newly-formed government retains the colonial military structures, regardless of the external threat environment.

This theory highlights a few important things about PGMs. First, it illustrates the ubiquity of *potential* pro-government militias. PGMs are relatively easy to establish and many times, the members of a PGM have a history of violence and are already connected to the networks that give rise to paramilitaries. This research also illustrates how governments connect with these networks, further illuminating the process of PGM formation and mobilization.

Mazzei's account of PGM formation is significantly more complex. Focusing specifically on Mexico, Colombia, and El Salvador, Mazzei argues that three factors lead to PGM formation. First, the country must undergo a structural change that threatens the existing political or social hierarchy, for example, democratization. This transformation splits the government between modernizers and conservative hardliners. When modernizers take power, hardliners seek recourse in order to protect the existing social order. Second, PGM formation requires that a segment of the population sees the existing order as legitimate and worth defending. Finally, just as with Ahram's analysis, there must exist networks that allow the government hardliners, to recruit and form a paramilitary. The PGM is created with political cover from politicians, financing from economic elites, and weapons and training from the state security forces.

This analysis introduces another useful variable when considering PGM formation: regime type. As mentioned in the previous review of research on paramilitaries, scholars have connected regime instability to increased PGM usage (Böhmelt and Clayton 2018) and have made arguments about internal violence and accountability (Carey, Colaresi, and Mitchell 2015). Mazzei's study also highlights the various links that PGMs may have to the government. In many cases, the central government is not directly involved with a paramilitary, rather a minister or group of government officials direct the PGM, creating potential problems for estimating the exact purpose of the PGM.

These two analyses focus more narrowly on state-initiated PGMs and what Carey et al. (2012) define as a "semi-official" PGM rather than an "informal" one. The state is much more involved in training, equipping, and funding semi-official PGMs and exerts much tighter control over their activities, as mentioned in the previous section. The following section will explore civilian-initiated groups which tend to be informal, especially at conception.[5]

1.4.2 Agentic Explanations of PGM Mobilization

Just as with the previous set of analyses, the overwhelming majority of explanations of community-driven PGM creation are qualitative studies of single countries. The distinguishing facet of this line of research argues that civilians are not passive victims of conflicts, but actively respond and shape conflict zones. When the risks of passivity grow too costly, civilians may become active participants in the conflict.

A group of these analyses focus on civilian defense forces (CDFs)—a particular type of PGM. Clayton and Thompson (2016) enumerate three criteria to distinguish CDFs from other types of groups. First, CDFs are composed primarily of civilians rather than former soldiers or rebels. Second, these groups operate in the area from which they were recruited. Finally, CDFs perform exclusively defensive tasks.

Drawing from field work in Mozambique, one analysis found that a military stalemate between the state and insurgent group—in this case, Renamo—resulted in high levels of civilian victimization, driving locals to form a militia to defend their community (Jentzsch 2014). Similarly, in Sudan, insurgent attacks against a community lead to the formation of a CDF (Blocq 2012). Insurgent violence is not the only factor that leads to CDF formation, however. Schubiger (2021) found that in the Peruvian civil war, high levels of civilian victimization *by the state* led civilians to create a paramilitary group to signal loyalty to the state and to avoid future victimization (Schubiger 2021). When the costs of a conflict rise, civilians have been shown to form militias to protect themselves both directly through active defense and indirectly through signaling. The effectiveness of this strategy, however, is disputed since the creation of a PGM can lead to more indiscriminate insurgent violence against the community (Clayton and Thompson 2016). This is an under-explored impact of PGMs on civil conflict dynamics.

---

[5] Carey et al. (2012) elaborate this point, noting that government-PGM relationships change over time. Although a PGM may form independently, the government may establish close ties with the paramilitary group, transforming it into a semi-official PGM.

Another group of studies on PGM formation focuses on ethnic defection. Despite expectations of group solidarity, especially in ethnic conflicts, studies have found that civilians are much less motivated by group identity than other factors. During the Algerian independence war and the Mau Mau insurgency in Kenya many Muslim Algerians and Kikuyu (the ethnicity associated with the Mau Mau movement) cooperated with the French and British authorities (Kalyvas 2008). Various incentives drove this process of ethnic defection. In the case of Algeria and Kenya, some coethnics feared retribution from the clearly predominant state. Others were motivated by local dynamics, material incentives offered by the state, and coercion. Finally, many people sought revenge against insurgent violence. This same pattern of intra-ethnic violence played out in Sri Lanka during the conflict with the Liberation Tigers of Tamil Eelam (LTTE). "Insurgent fratricide" compelled certain groups of Tamils to collaborate with the state to gain protection from rival Tamil leaders in the LTTE (Staniland 2012).

Finally, recruitment patterns may also reveal the incentive structure that leads to PGM formation. Studies have found that PGM members have a similar profile to insurgents. Both counterinsurgents and rebels tend to be from poor, rural, and politically marginalized groups and communities (Humphreys and Weinstein 2008). People may join a PGM to secure material benefits from the group, mirroring greed-type explanations of insurgency (Collier and Hoeffler 2004). Drawing from a study of the Colombian civil war, the main factor that determines whether a person will join an insurgent group or PGM seems to be who controls the recruit's community. Where the state exercises sovereignty, people are more likely to join a PGM. Where insurgents predominate, recruits join the insurgency (Arjona and Kalyvas 2008). Private agendas may also induce communities to form a PGM to gain an advantage over a rival community (Peic 2019).

We can generalize this motivation to anticipate that certain PGMs may form for material incentives from the state. Otherwise, endogenous conflict dynamics such as fratricide or intra-communal competition may compel otherwise peaceful communities to arm and fight alongside the state. However, no one has yet published a large-N quantitative analysis of PGM mobilization. I draw on these theories to form hypotheses in the following chapter. This short review of the literature highlights the variety of factors which drive PGM mobilization and illustrates how the current analysis fits into existing work on pro-government militias.

Chapter 2

DATA, MODELING STRATEGIES, AND PRELIMINARY HYPOTHESIS TESTS

As can be seen from the previous chapter, numerous theories account for PGM formation in the literature. Within the context of ongoing intrastate conflict, however, existing theories can be simplified into two theoretical expectations: PGMs as force multipliers and PGMs as intelligence gatherers.

Numerous works emphasize the military functions of PGMS. Biberman (2018) argues that states must exchange strict control over PGMs in order to recruit more militarily effective militias. In Myanmar, pro-government militias were the primary counterinsurgent force as the central military underwent a transformation in the 1990s (Eck 2015). PGMs that share an ethnicity with the host state better augment the state's military capability and lead to higher levels of violence (Abbs, Clayton, and Thomson 2020). Clayton and Thompson (2014) show that when a PGM joins a conflict, the PGM itself becomes the focal point of insurgent violence, leading to more intense fighting and higher casualties. Both Jentzsch, Kalyvas, and Schubiger's (2015) and Carey and Mitchell's (2017) survey of existing literature focus almost exclusively on the violent functions of PGMs in civil conflict, supporting the idea that PGMs are primarily violent actors, utilized as an auxiliary counterinsurgent force.

Examples of violent, counterinsurgent PGMs abound. The Myanmar government successfully recruited the United Wa State Army to combat numerous insurgencies in the golden triangle region of Myanmar. The Serbian government equipped and trained the Arakan Tigers to target and massacre Bosnians during the Yugoslav War. And as mentioned previously, the Iraqi government supported the Kurdish Peshmerga to combat the Islamic State. These brief examples clearly illustrate the various violent forms that PGMs can take.

Conversely, other works highlight the intelligence gathering functions of PGMs. Kalyvas (2006) boldly asserts that pro-government militias are "primarily a political rather than a military institution," (107) used to extend the host state's local control and effectively monitor a territory. In many cases, states rely on local informants to overcome the "identification problem" (Kalyvas 2006, 89), allowing the state to selectively target insurgents and their supporters. Within a counterinsurgency campaign, gaining access to reliable intelligence is paramount for success (Army 2014), and numerous articles emphasize how a militia's superior local knowledge provides this information (Clayton and Thompson 2016; Peic 2014; Lyall 2010; Carey and Mitchell 2017).

To illustrate, the Naparama militia in Mozambique served as a primarily defensive force, only resorting to arms when absolutely necessary and primarily alerting the government ot the movements of Renamo (Jentzsch 2014). During the second Chechen War, Chechen militia groups would cooperate with Russian forces to identify and apprehend rebels and to levy credible threats against potential supporters. Critically, however, their involvement in the conflict resulted in fewer fatalities and reprisal attacks, highlighting their value as an intelligence group rather than a violent group (Lyall 2010). Lastly, Staniland emphasizes how the Sri Lankan government recruited competing Tamil factions to get intelligence on the LTTE in exchange for protecting Tamil defectors (Staniland 2012). It is clear that while many PGMs engage in anti-insurgent violence, one of their primary tasks is to relay quality information to the state.

Although these two hypotheses are not necessarily competing—PGMs could be used both for their local knowledge and for their military capabilities against insurgents—these two broad theories provide a clear opportunity to quantitatively explore when states will utilize PGMs. Up until this point, no other scholar has published a large-N analysis on PGM mobilization in civil conflict, leaving a gap in the literature. I generate two hypotheses to explore the context in which PGMs join and participate in a civil conflict. To test the force multiplier theory, I focus on the state's need for military assistance since greater military threat should drive the state to find ways to balance against the threats it faces.

> H_1: Where the state faces greater military threat, it will mobilize PGMs as a force multiplier.

Similarly, the intelligence gathering hypothesis relies on the states need for information about NAG operations and sympathizers.

> H_2: Where the state cannot effectively gather intelligence on insurgent groups, it will mobilize PGMs.

Although I juxtapose these hypotheses for the purposes of this analysis, they are not mutually exclusive. A PGM may be used both for its military capabilities against NAGs and for its ability to relay quality local information to the state. In line the nature of theses theories, I expect to find support for both hypotheses in my quantitative tests. Comparatively, however, I expect the intelligence gathering hypothesis to better explain PGM participation in civil conflict based on PGMs' local character and general lack of training compared to the state's military.

The following section will outline the data-gathering process for this analysis. I will begin by documenting the source of each variable, the theoretical justification for inclusion, and summary information for each. I will enumerate important coding decisions, transformations and adjustments made to the final dataset to improve model performance. I will then discuss my modeling choices and report the results of the preliminary regressions.

Before continuing to the data collection process and hypothesis tests, however, I must clarify two points. First, this analysis focuses on PGM *mobilization* rather than *formation*. The existing literature on PGMs still has not adequately studied the process of paramilitary formation. Thus, in the Pro-Government Militias Dataset (PGMD) Carey, Mitchell, and Lowe 2012, groups "form" at the first mention of activity on the side of the government, not necessarily when the group actually came to be. Second, in this analysis and discussion, I do not distinguish between state-initiated and civilian-initiated groups. I focus primarily on the state's decision to implement or mobilize PGMs since explicit connection to the state is required for inclusion as a PGM. The state must at least accept the PGM's support, regardless of whether the state initiated or materially supports the group.

2.1 Data Source and Summary

For all models, the unit of analysis is country conflict-year and the population of cases are drawn from the Uppsala Conflict Data Program (UCDP) dyadic dataset (Harbom, Melander, and Wallensteen 2008). I narrow the entire dataset to intrastate conflict years and then aggregate them to country conflict-year. Observations begin in 1946 and continue until 2007 when the pro-government militia database stops coding new PGMs. This version of UCDP data was chosen to aid in calculating the number of active NAGs in a given year as well as whether the host country was involved in an interstate conflict in a given year. The final population of cases includes 1,243 civil conflict country-years across 93 countries.

2.1.1 The Force-Multiplier Theory

In order to test the force-multiplier hypothesis, I collected numerous explanatory variables. Five variables are derived from the UCDP dyadic dataset:

- *N.NAGs* provides the number of active non-state armed groups in a given country-year. In line with UCDP coding rules, an insurgent group is coded as active if the conflict causes more than twenty-five conflict causalities in a given year. Multiple insurgent groups should be more difficult for governments to repress combat effectively, driving the state to utilize PGMs to enhance their military capabilities.
- *InterstateConf.* is a count variable denoting the number of interstate conflicts in which the host country was a main actor (coded as Side A or Side B in the UCDP dataset) in a given year. If the host state is involved in an interstate conflict, the competing demands on its military forces should push the state to reserve its conventional forces for the interstate conflict and to rely PGMs

to combat domestic insurgencies in line with Eck's (2015) study. Six percent of observations include at least one interstate conflict.

- *ForeignCommits* is another count variable that reflects the number of active foreign conflicts in which the host state has committed troops.[1] Following the same logic as the previous variable, foreign troop commitments reduce the state's ability to combat an insurgency. Seven percent of observations include at least one foreign troop commitment.
- *N.StateSups* is a count variable that reflects the number of foreign states that have sent troops to assist the host state. This variable does not capture other forms of support such as material or political aid. Where the host state has external supporters, it should be less likely to use PGMs to combat a NAG. Thirteen percent of observations include at least one external state supporter for the host state.
- *N.RebelSups* is the same conceptualization and theoretical justification as *N.StateSups* but for the NAG. Five percent of observations include at least one external state supporter for the NAG.

I include a measure of NAG strength since stronger groups should induce a state to search for a force multiplier. I use Cunningham, Gleditsch, and Salehyan's (2009) Non-state Actor Data (NSA). This dataset was built on the UCDP dataset in 2008 and codes numerous variables relating to NAG size and fighting capabilities. The variable I use—*Reb.Strength*—is a composite measure that reflects the insurgents' mobilization, arms procurement, and fighting capacity. It is ordinal from one to five, with one representing complete state preponderance, three representing parity between insurgents and the state, and five representing insurgent preponderance. On account of updates to the UCDP dataset since then, I was unable to consistently match insurgent groups between the different versions. Therefore, I matched country-years between the my dataset and the NSA. Where multiple active insurgencies are present in a given observation year, I took the maximum value from

[1]This variable reflects whether the host state is coded as a secondary actor in another conflict in the full UCDP dataset

Cunningham *et al.'s* rebel strength variable. I chose to take the maximum rather than the mean on the assumption that the datasets are more likely to disagree on minor insurgencies rather than those that are highly organized and pose a significant challenge to the host state. The NSA data covers years from 1946 to 2002, resulting in missingness in six percent of the observations. Since most of this missingness is between 2002 and 2007, I impute these missing observations with the most recent coded value, making the assumption that rebel strength does not change significantly between 2002 and 2007. The NSA does not offer an estimate of uncertainty for the rebel strength variable.

N.Rivals reflects the number of the host state's external rivals since an ongoing rivalry may induce it to reserve its conventional military forces in anticipation of an interstate conflict and rely on PGMs to repress internal threats. I use Thompson and Dreyer's rivalry data for this variable (Thompson and Dreyer 2011).

I also collected a measure of peripherality to account for the diminishing power of the state's military over distance. If a group is located farther from a state's center, the government should have a harder time combating the NAG. For theoretical reasons, I don't use this variable in the main regressions

Following the example of the existing literature on peripherality, I use the distance to the capital to estimate this variable. On account of incomplete data, I utilize Buhaug *et al.'s* (2009) replication dataset for all observations before 1989. For observations in 1989 and later, I combine data from UCDP Georeferenced Events Data (GED) and PRIO GRID (Sundberg and Melander 2013; Tollefsen, Strand, and Halvard 2012). The GED is UCDP's most disaggregated dataset, providing latitude and longitude of each violent event in a conflict. It also includes the corresponding ID for PRIO GRID units. PRIO GRID utilizes a geographic reference system that divides all territory on earth into exclusive grid squares, no larger than 3,025 km^2. The dataset includes a measure of the distance from the center of each grid square to the capital of the host country, providing geographical precision of no more than 39 kilometers for all violent events.

I drop multiple yearly observations of violent events in the same GRID unit from the GED. I do this because multiple violent events in the same location likely reflects an area where the government can militarily engage with the NAG. Thus, each grid square hosting a violent event in a year is given equal weight when calculating the mean distance to the capital. I take the natural logarithm of this variable to correct the uneven distribution of values. Since some observations include violent events in the capital—distance to the capital is zero—I add 0.01 to all values to prevent errors when logging the variable. I scale the logged values to improve the model's efficiency.

Finally, I include a measure of military spending taken from the World Bank (World Bank) and downloaded through the `WDI` API (Arel-Bundock) and the number of military personnel taken from the Correlates of War National Military Capabilities dataset (Singer, Bremer, and Stuckey 1972). Higher values of both military spending and military personnel should decrease the chances of a state resorting to a PGM. I log and scale both variables.

To test the force multiplier hypothesis I utilize a principal component to estimate the joint effect of most these variables. Traditionally principal component analysis (PCA) is used as a dimensionality reduction tool to estimate the combined effect of numerous, correlated control variables. In this analysis, however, I use it as the main explanatory variable to measure the latent military threat facing the state. This decision presents problems for direct interpretation since the values of the PC do not reflect observed values, but we may still interpret the sign and magnitude of the final estimate to gauge the strength of the force multiplier hypothesis. The R package I used to calculate the PCA—*pcaMethods*— does not provide an estimation of uncertainty and has assumptions of linearity standard in principal component analysis.

I use only eight of the previously mentioned variable to generate the PC: interstate conflict, foreign troop commitments, state supporters, rebel supporters, rebel strength, state rivals, military spending, and military personnel. Despite their theoretical support for the force multiplier hypothesis, *N.NAGs* and *Cap.Dist.Log* could also reflect problems monitoring insurgencies. Each additional NAG in a state's territory requires dedicated personnel and operations, potentially outstripping the state's intelligence apparatus. States are less likely to effectively penetrate peripheral areas, further complicating intelligence gathering when a NAG operates in these areas. Thus, where there are multiple insurgent groups and groups are farther from the state's capital, states should be more likely to rely on locally-based PGMs to provide information on NAGs.

Since these two variables could support both theories, I do not use them in the initial hypothesis tests. I explore their effect in Chapter 3 where I interact them with my primary explanatory variables to determine whether they support one theory more than the other.

Finally, to overcome the missingness that usually impedes principal component analysis, I use the pcaMethods package (Stacklies et al. 2007) in `R` which enables me to model the missingness as a neural network and impute the missing data points. This package does not provide an estimate of uncertainty, however. This maximizes the number of observations I can use and ensures that I have a PC value for each observation. Lastly, since the PC accounts for variation among closely correlated variables, I check the direction by comparing the correlation of the PC with values of *Reb.Strength*. I ensure that larger values of the PC correlate with stronger rebel groups, thereby reflecting greater military threat to the state.

2.1.2 The Intelligence Gathering Theory

To proxy the state's need for intelligence on NAGs, I use the Historical Index of Ethnic Fractionalization (HIEF) (Drazanova 2019). This variable follows Fearon's (2003) original coding scheme whereby the country-year value is the probability that two randomly selected citizens are from different ethnic groups. The HIEF simply extends this variable to account for change over time. Although this measure has been criticized for failing to accurately capture the actual causal mechanisms that lead from ethnic difference to conflict (Cederman and Girardin 2007), in this analysis, I am primarily interested in the difficulty a state may face in gathering intelligence and monitoring its territory. The proposed causal mechanism is simply that linguistic and ethnic diversity in a society will raise the costs of gathering intelligence on the activity and movements of an insurgent group.

The U.S. Army counterinsurgency field manual highlights the critical role of local culture in shaping intelligence gathering strategies. "Understanding the beliefs and values of a local population is critical for effective information operations. Failure to respect or understand the beliefs of a local population can result in serious hostility towards military intervention and attacks upon military personnel" (Army 2014, 3-14). Where a counterinsurgent force operates in a highly diverse and ethnically fractionalized society, it will surely incur greater costs in navigating the different cultures and gathering intelligence.

Furthermore, different ethnic groups may hold distinct priorities within the conflict. For instance, intercommunity rivalry may drive one group to provide false information to the state to settle a local dispute as illustrated by Blocq 2012 in his analysis of militia formation in South Sudan. In this specific case, the local community was much more concerned with what they perceived as hostility from a rival community rather than defeating insurgents. Examples of civilians providing false information to settle personal disputes abound in the literature on civil wars (Kalyvas 2006). Higher levels of ethnic diversity surely increases the chance of these disputes clouding the state's attempt to gather reliable intelligence on the insurgency. Mistrust also characterizes conflict zones as civilians fear being "denounced" (Kalyvas 2006, 178) by their neighbors. Ethnic diversity could easily become a focal point of this mistrust, both increasing polarization (Kalyvas 2006, 181) and leading to false information.

As can be seen, higher levels of ethnic fractionalization can seriously complicate a state's intelligence gathering efforts, driving states to resort to local actors or militias to relay reliable intelligence. Although ethnic fractionalization is an imperfect measure, it accounts for much of the variation that a more sensitive independent variable would capture. For instance, inter-group distrust or civilian support for the insurgency may be the explicit, micro-level causal mechanism that frustrates the state's intelligence gathering efforts but testing these mechanisms would require significantly finer-grained data which currently does not exist. Ethnic fractionalization serves as a strong proxy variable for the difficulty a state faces when gathering intelligence on its domestic insurgencies.

Table 2.1 shows summary statistics for all explanatory variables in the analysis.

Table 2.1: Summary statistics for all variables.

	Min.	Mean	Median	Max.	Variable Type
N.NAGs	1	1.75	1	10	Count
InterstateConf.	0	0.07	0	3	Count
ForeignCommits	0	0.1	0	5	Count
N.StateSups	0	0.7	0	96	Count
N.RebelSups	0	0.1	0	30	Count
Reb.Strength	1	1.9	2	5	Ordinal
Cap.Dist.Log	0	497	331	3,360	Continuous (kilometers)
N.Rivals	0	1.4	1	6	Count
Mil.Spend	1.04	5,984.6	429.6	589,586	Continuous (current USD in millions)
Mil.Per.	1	229	85	4,300	Continuous (in thousands)
EthnicFrac	0.01	0.54	0.57	0.89	Index

2.2 Dependent Variables, Control Variables, and Modeling Choices

I test both hypotheses with two different modeling strategies. I draw both dependent variables from the Pro-Government Militias Database (PGMD) (Carey, Mitchell, and Lowe 2012). This data provides information on over 300 PGMs including the group's name, recruitment base, connection to the government, and active years.

The first model seeks to address when states will mobilize PGMs, so I use a hazard model. The Cox proportional hazard model is improper in this circumstance, however, on account of its inability to deal with unbalanced panel data, multiple failure events, and time-varying covariates. Obviously in attempting to model multiple PGM formations using dynamic explanatory variables, these limitations present major challenges for this analysis. Therefore I use Andersen and Gill's extension of the Cox proportional hazard model which integrates a multivariate counting process into the Cox framework. Instead of simply calculating changing hazard rates for time-invariate covariates, this model accounts for multiple failure events, taking previous failure events and dynamic explanatory variables into account when calculating the variable hazard rate. A full specification of the model is provided below:

$$\lambda_i^{(n)}(t) = Y_i^{(n)}(t)\lambda_0(t)\exp\{\beta_0' Z_i^{(n)}(t)\}$$

Where λ is the intensity, rather than hazard rate, for each individual i at time t. Y is a dummy variable to indicate when the ith individual is under observation. λ_0 is the baseline hazard. And finally, Z is a matrix of covariates.

The Andersen-Gill extension is a better option for modeling PGM formation since many countries will mobilize more than one PGM and conflict characteristics will change over time, generating different hazards and intensities at different times. This model will also account for autocorrelation in the hazard rate after an initial PGM formation since once a state begins using PGMs, it may be more likely to utilize PGMs in the future. The main drawback of the Andersen-Gill model is its inability to account for numerous paramilitary mobilizations in the same year. To account for this, I transform the number of PGMs mobilized in a country-year into a binary measure of whether a new PGM began activity.

To overcome the problem of unbalanced data, I artificially balance my dataset to ensure that all observations are yearly. A country enters the dataset when it is first included in the UCDP dyadic dataset. The state has yearly observations until the last year the it is included in the UCDP dataset. To illustrate, Spain has an intrastate conflict episode from 1978 to 1982 and another one from 1985 to 1987. The artificial balancing process adds observations for Spain from 1982 to 1985 so that Spain is represented continuously between 1982 and 1987. This process adds 880 non-conflict country-years, but it is the most principled way to ensure that the final estimates aren't biased. Paramilitaries formed in 190 country-years in the balanced dataset. In forty observations, more than one PGM formed. 80% of PGM mobilizations occur during active conflict years.

The second model seeks to determine when states will utilize multiple PGMs. I implement a negative binomial regression model to estimate how my chosen covariates influence the number of active PGMs in a given country-year. Since balancing is unnecessary for this model, my population of cases only includes intrastate conflict years. 151 PGMs form in this dataset, and at least one PGM is active in 61% of observation years. I fit a temporal spline for each country to account for autocorrelation in the outcome variable, since a PGM will persist for multiple years after formation. More agnostic strategies exist to account for this autocorrelation, as illustrated in Carter and Signorino (2010) and in Rasmuessen and Williams (2006), but I choose to use a temporal spline for its simplicity and more general understandability. Obviously, the negative binomial model reflects a different data-generating process than the hazard, but taken together, the hazard and negative binomial models should provide a clearer understanding of PGMs' role in civil conflict and why states will use one or many PGMs throughout a conflict.

Unless otherwise noted, I include two control variables in every model. First, I control for GDP/capita (World Bank) since lower values have been linked to increased internal conflict and higher rates of both NAG and PGM recruitment. Furthermore, poorer states may be unable to support a strong, central military and may rely on PGMs to act as the government's coercive apparatus. I also control for regime type with V-Dem's liberal democracy index (Coppedge, Gerring, et al. 2021). Weak democracies are more likely to have internal conflict and, as detailed in the previous chapter, are more likely to utilize PGMs to avoid international accountability (Carey, Colaresi, and Mitchell 2015).

2.3 Results and Discussion

In the hazard model, the PC returned positive results revealing that higher levels of latent military threat increase the chance of a state utilizing a PGM, but the result was insignificant, casting doubt on the robustness of this relationship. Although interpreting control variables is improper, regime type was positive yet also insignificant, aligning with Carey et al.'s (2015) expectation that weak democracies will be more likely to use PGMs. GDP/capita's coefficient was negative, suggesting that poorer countries are more likely to use PGMs; however, this results also failed to reach traditional levels of significance.

Ethnic fractionalization, on the other hand, better predicts PGM mobilization. I ran two models to test its effect. The first model is a simple bivariate test, justified by the fact that both control variables have disputed causal relationships with ethnic fractionalization. The second model, includes controls. Regardless, ethnic fractionalization maintains similar magnitudes and significance levels. Results can be seen in Figure 2.1 and Table 2.2. These results lends strong support to the intelligence gathering hypothesis and casts doubt on the force multiplier theory. In general states that are highly fractionalized will be more likely to mobilize a PGM, but the level of internal military threat seems to have negligible effects on the decision to utilize a PGM.

The results of the negative binomial models largely agreed with those from the hazard models. Once again, ethnic fractionalization significantly outperformed the PCA. Both main explanatory variables maintained the same sign and significance levels. Interestingly, in the force multiplier model, V-Dem and GDP/capita were significant with the expected signs. In the ethnic fractionalization model, GDP/capita changes signs, yet is insignificant. This is likely due to the amount of variation accounted for by *EthnicFrac*. Since none of these models are designed to test explicitly the impact of V-Dem or GDP/capita on the number of active PGMs, however, one cannot make conclusive inferences on these variables. Figure 2.2 and Table 2.3 report the results of these models.

Taken together, we can tentatively conclude that highly fractionalized states are both more likely to implement a PGM *and* more likely to use multiple PGMs. I argue that this is chiefly due to the difficulty that these states have gathering intelligence on domestic insurgencies. The null results drawn from the force multiplier models suggest that when a state faces militarily stronger opponents, they do not consistently turn to PGMs as force mulitpliers.

In the aggregate, PGMs are not used for their violent capabilities against in-

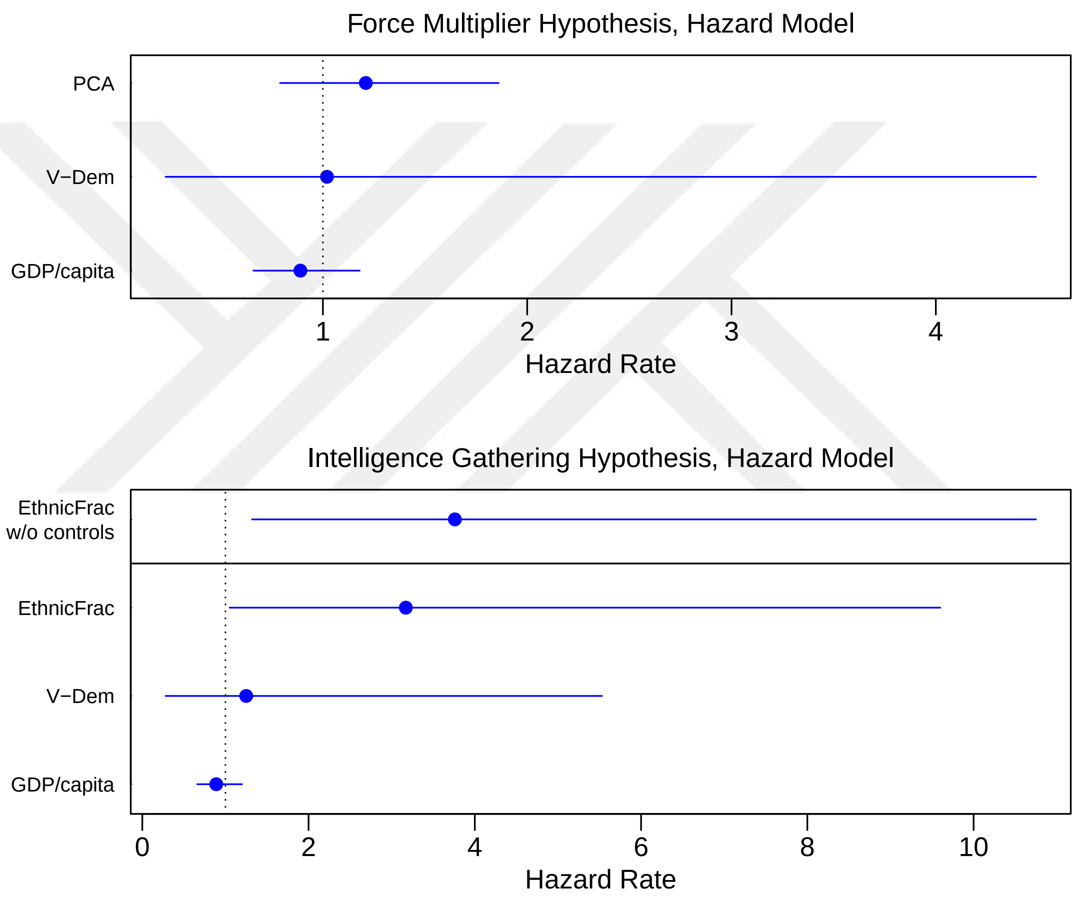

Figure 2.1: Hazard rate plots for both hypotheses, Andersen-Gill Hazard Model.

	(1)	(2)	(3)
PCA	1.213		
	(0.217)		
EthnicFrac		3.759**	3.168**
		(0.536)	(0.566)
V-Dem	1.022		1.248
	(0.755)		(0.760)
GDP/capita	0.886		0.891
	(0.147)		(0.151)
Observations	1,842	2,223	1,723
Penalized Log-Likelihood	−513.25	−583.07	−489.35
θ	0.920*** (0.296)	0.805*** (0.268)	0.781*** (0.275)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2.2: Hazard Rates for Andersen-Gill Models

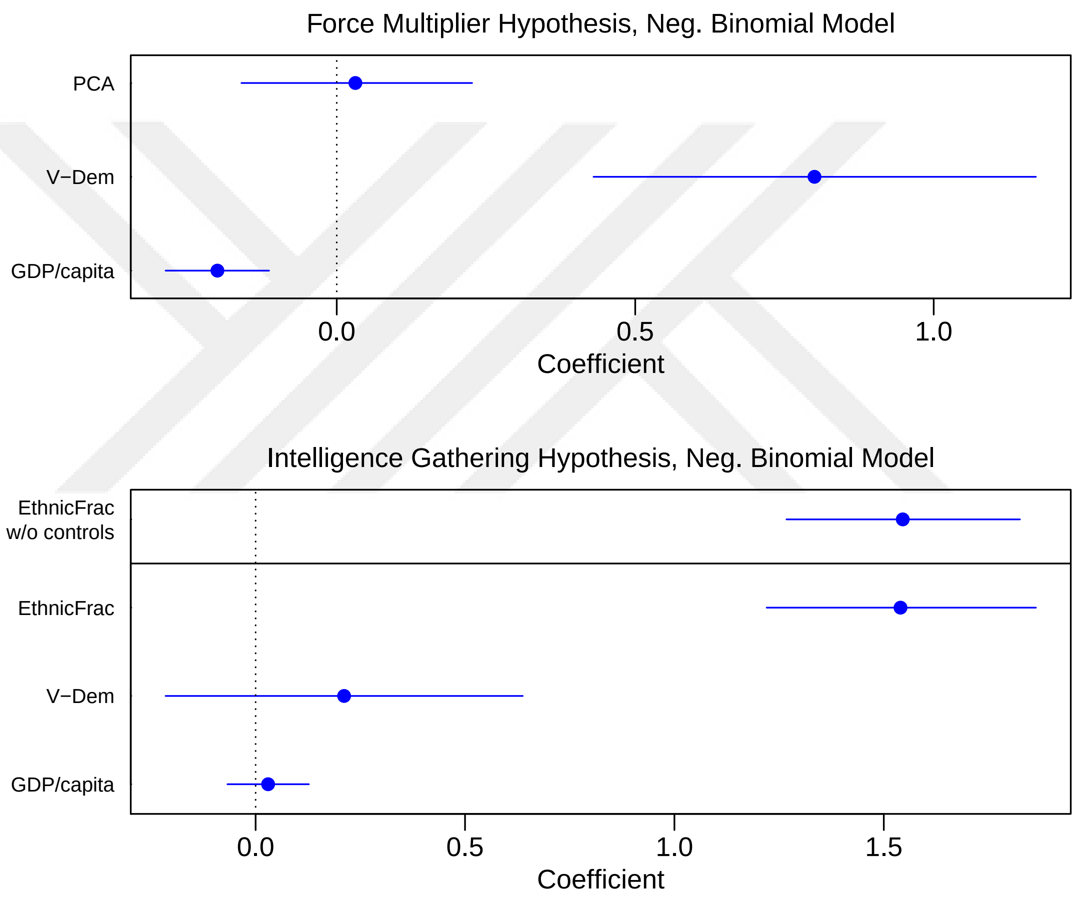

Figure 2.2: Coefficient plots for both hypotheses, negative binomial model.

Table 2.3: Coefficients for Negative Binomial Models

	(1)	(2)	(3)
PCA	0.031		
	(0.099)		
EthnicFrac		1.545^{***}	1.540^{***}
		(0.144)	(0.168)
V-Dem	0.800^{***}		0.211
	(0.189)		(0.213)
GDP/capita	-0.200^{***}		0.030
	(0.043)		(0.047)
Constant	-1.375^{***}	-2.567^{***}	-2.309^{***}
	(0.110)	(0.135)	(0.160)
Observations	962	1,161	911
Log Likelihood	−1,602.488	−1,610.444	−1,450.264
θ	2.432^{***} (0.223)	3.199^{***} (0.320)	2.854^{***} (0.287)
Akaike Inf. Crit.	3,218.977	3,230.888	2,914.529

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

surgents but for their intimate local knowledge. These results clarify the role that PGMs play within civil conflict. An undue emphasis on PGMs as counterinsurgents risks misrepresenting their actual purpose. Of course, in some cases, militias are responsible for extreme or indiscriminate violence, but in general, they are not violence specialists. These results provide guidance for future research on paramilitaries in civil conflict, as well. The current literature on PGMs primarily focuses paramilitary violence, but if, as this analysis suggests, violence is an auxiliary goal or even byproduct of PGM involvement in civil conflict, research should reflect this hierarchy. I will extrapolate these points further in chapter four.

To further examine this conclusion, however, I will conduct some exploratory tests concerning the force multiplier hypothesis to determine whether some adjustments may lead to a stronger statistical relationship between military threat and PGM participation in civil conflict.

Chapter 3

FURTHER EMPIRICAL TESTS

Thus far, it seems that ethnic fractionalization and the need to gather intelligence on insurgent groups best explains PGM participation in civil conflict. This chapter further explores the comparative explanatory power of both hypotheses. First I test which combination of force multiplier variables best predict PGM mobilization and participation in civil conflict. Can a more parsimonious collection of these variables provide more explanatory leverage than the eight that were included in the previous chapter's principal component? I will then turn my attention to the number of NAGs and the peripherality of insurgent groups. Both variables exhibit a strong positive correlation with PGM mobilization and could be used to justify either theory, but do these relationships reflect one theory more than the other? Finally, I will build a full predictive model for both outcomes to determine which combination of variables best predicts when and how many PGMs mobilize in a conflict. This analysis will clarify whether one theory best accounts for PGM involvement in civil conflict or if both theories explain part of the causal story. The results of this test will provide guidance for future research on PGM formation and will further illuminate the process PGM mobilization in civil conflicts.

3.1 Exploring the Force Multiplier Variables

In the previous models, I utilized eight variables in the PCA to proxy the latent military threat to the government. A brief look at the bivariate relationships (Figure 3.1) between each of these variables and both outcomes shows that most of them fail to reach statistical significance, potentially generating variation in the principal component that does not help explain variation in the outcome.

In order to test whether some combination of these eight variables better explains

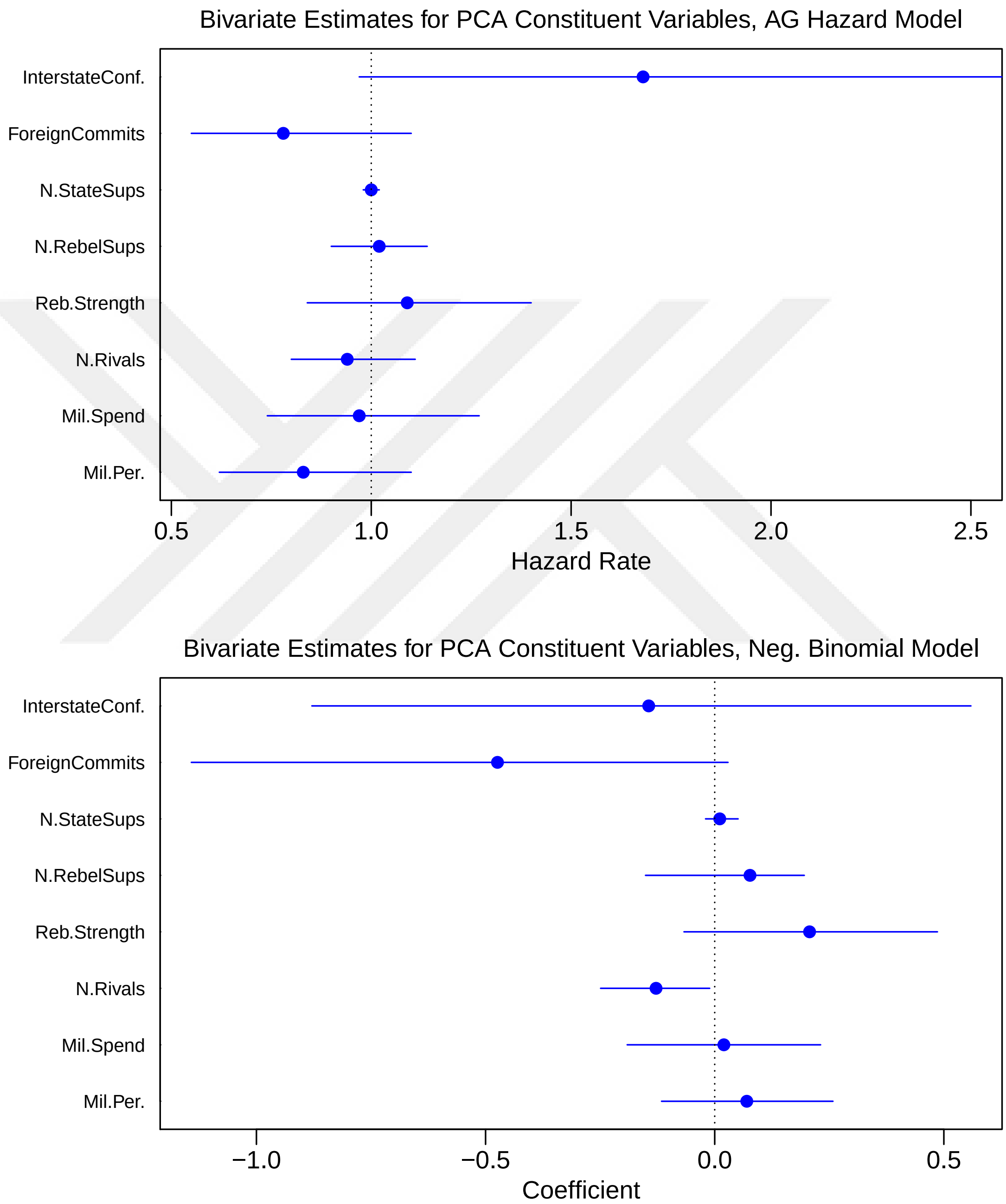

Figure 3.1: Bivariate estimates for force multiplier variables.

PGM participation in civil conflict, I rely on out-of-sample predictive tests. First, I divide the dataset into a 80% training and 20% test sample. Then I run the hazard and negative binomial tests on the training samples to generate estimated regression coefficients. I use these parameters to generate predicted outcomes for the test set. I judge the accuracy of these predictions by calculating the mean squared error for each different combination of variables. As shown in Equation 3.1, the mean squared error takes the average of the squared difference between the actual (Y_i) and predicted ($\hat{Y}_i$) outcomes. Lower values represent higher predictive accuracy. To ensure the robustness of my results, I repeat these calculations on 500 different samples and average the results across the iterations.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{3.1}$$

The out-of-sample test results for the AG hazard model suggest that *Reb.Strength* alone best predicts PGM mobilization. In the negative binomial model, the optimized combination of variables is the number of foreign troop commitments, state supporters, rebel supporters, and state rivals. I created a new, optimized PC that includes only these four variables. I re-run both models using rebel group strength as the primary explanatory variable in the AG hazard model and the optimized PCA in the negative binomial model. Just as in the previous chapter's models, I include V-Dem's liberal democracy index and GDP/capita as control variables in both regressions. In the negative binomial model, I also include the same temporal spline to account for autocorrelation in the outcome. The results of these models are reported in Figure 3.2.

As can be seen, *Reb.Strength* is positive yet insignificant. Considering the results of the previous PCA and the bivariate relationship, this result is expected. In the negative binomial model, however, the optimized PC exhibits a much stronger relationship, reaching traditional levels of significance. Initially, this improved statistical relationship seems to indicate that the data supports the force multiplier hypothesis. However, a quick comparison of the log-likelihood between the optimized PCA and ethnic fractionalization reveals that despite the stronger statistical

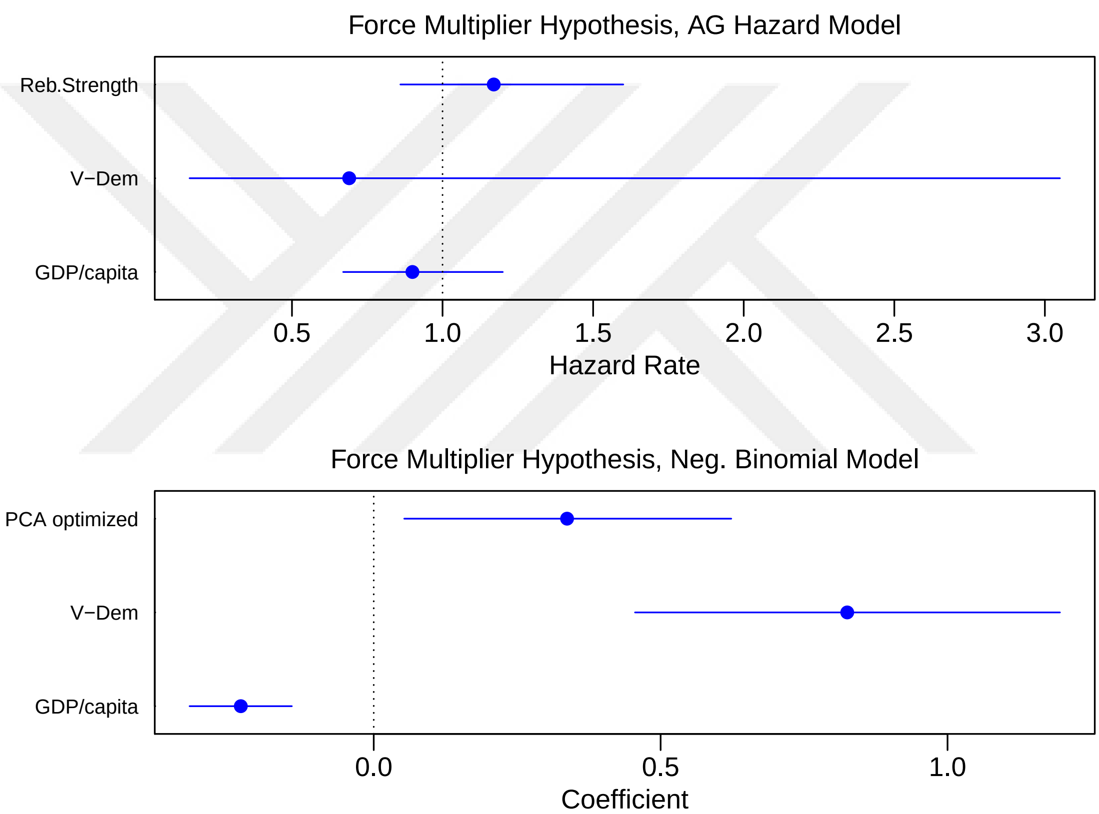

Figure 3.2: Force multiplier models with optimized variable selection.

relationship, the *EthnicFrac* models account for more variation. This result confirms the previous chapter's conclusion that states resort to PGMs when the state needs quality intelligence, rather than when they face greater military threat. Table 3.1 reports regression coefficients and log-likelihoods of both models.

	AG Hazard		Negative Binomial	
	(1)	(2)	(3)	(4)
PCA Optimized			0.741***	
			(0.141)	
Reb.Strength	0.199			
	(0.164)			
EthnicFrac		1.247**		1.540***
		(0.559)		(0.168)
V-Dem	−0.192	−0.212	0.078	0.211
	(0.809)	(0.777)	(0.226)	(0.213)
GDP/capita	−0.135	−0.072	−0.040	0.030
	(0.154)	(0.151)	(0.048)	(0.047)
Constant			−1.440***	−2.309***
			(0.124)	(0.160)
Observations	1625	1625	911	911
Log Likelihood	−471.74	−470.08	−1,483.414	−1,450.264
θ	0.832***	0.664***	2.313***	2.854***
	(0.282)	(0.253)	(0.210)	(0.287)

Note: The reported log-likelihood of the AG hazard models is penalized.
*p<0.1; **p<0.05; ***p<0.01

Table 3.1: Regressions using explanatory variables from PCA optimization tests.

3.2 *Exploring the Impact of Multiple NAGs and Distance to the Capital*

As mentioned in the previous chapter, *N.NAGs* and *Cap.Dist.Log* could support both the force multiplier and intelligence gathering hypothesis. A larger number of active insurgent groups may overstretch the state's military by fighting in different areas and coordinating attacks.[1] Additionally, power diminishes over distance, making peripheral insurgencies more difficult to effectively combat. On the other hand, multiple NAGs may strain the intelligence gathering capabilities of the state since each additional group requires dedicated personnel and concentrated effort to monitor. This same logic holds for gathering intelligence on rebels that are farther from the state capital. States are much less likely to have effectively penetrated peripheral territories, allowing NAGs to operate more freely.

So do these variables lend more support to one hypothesis or the other? In order to test this, first, I will examine the independent effect of both variables in their own models. Then, I build interaction models to track the impact of *N.NAGs* and *Cap.Dist.Log* conditional on the PCA and ethnic fractionalization. A highly conditional relationship would reflect support for the corresponding theory. For example, if the effect of *N.NAGs* is conditional on the PC, it suggests that the number of insurgent groups only increases the chance of PGM usage when military threat is high, lending support to the force mulitplier hypothesis. Finally, to estimate the strength of these interactive effects, I will use out-of-sample prediction. Improved predictive ability reflects the usefulness of the interacted effect for explaining the outcome and further confirms the variables support for the respective hypothesis.

3.2.1 Baseline Regressions

For the initial, baseline regressions, I include the same controls for V-Dem's liberal democracy scores and GDP/capita. Just as in previous models, I include a temporal cubic spline in the negative binomial model. When comparing models, I remove incomplete observations to prevent different sample sizes from impacting model fit

[1] Evidence of this type of coordination was shown in Akcinaroglu 2012.

estimators. In order to prevent negative values of *Cap.Dist.Log* from changing sign when multiplied with negative values from the PCA, I add the minimum value of *Cap.Dist.Log* to every observation. This procedure preserves the variation in the variable while ensuring that the interaction term behaves properly.

The results from the baseline models with *N.NAGs* and *Cap.Dist.Log* are unsurprising. In the AG hazard models, the number of insurgent groups is positive and significant while logged distance to the capital barely fails to reach traditional levels of significance. In the negative binomial models, both variables are positive and significant. These results can be seen in Figure 3.3. Table A1 in the appendix reports the regression results. In each case, more insurgent groups and peripheral insurgencies lead states to mobilize PGMs.

3.2.2 Interaction Models

In the Andersen-Gill models, the interacted effects present interesting yet mostly inconclusive results. *N.NAGs* exhibited a decreasing effect at higher levels of latent military threat and at the highest levels of ethnic fractionalization. This suggests that the number of insurgent groups may support the intelligence gathering hypothesis, but the relationship is rather weak. Logged distance to the capital has an increasingly positive effect on the decision to mobilize PGMs conditional on both military threat and ethnic fractionalization, but fails to reach standard significance levels. Based solely on the hazard models, *Cap.Dist.Log* does not conclusively support one hypothesis more than the other. The marginal effects plots for the hazard models can be seen in Figure 3.4. A table of the regression results is reported in Table A2 in the appendix.

In the negative binomial models, the effect of *N.NAGs* conditional on the PC once again is negative and insignificant at higher levels military threat. At the lowest values of ethnic fractionalization, however, the number of insurgent groups has a negative and significant effect. But as *EthnicFrac* increases, the impact of *N.NAGs* also increases, changing signs and becoming statistically significant at the highest levels of ethnic fractionalization. This suggests that the effect of *N.NAGs*

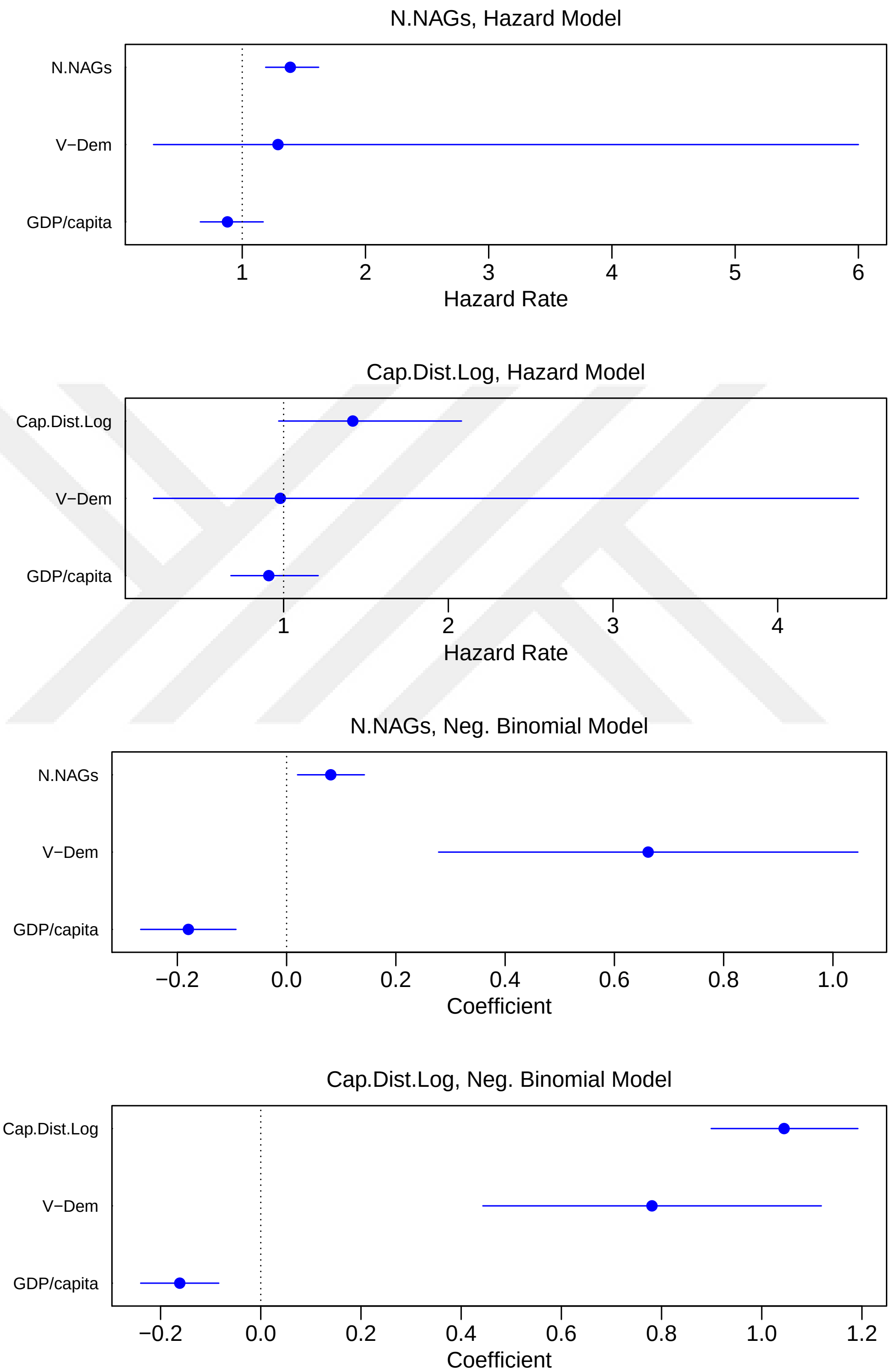

Figure 3.3: Base models for *N.NAGs* and *Cap.Dist.Log*.

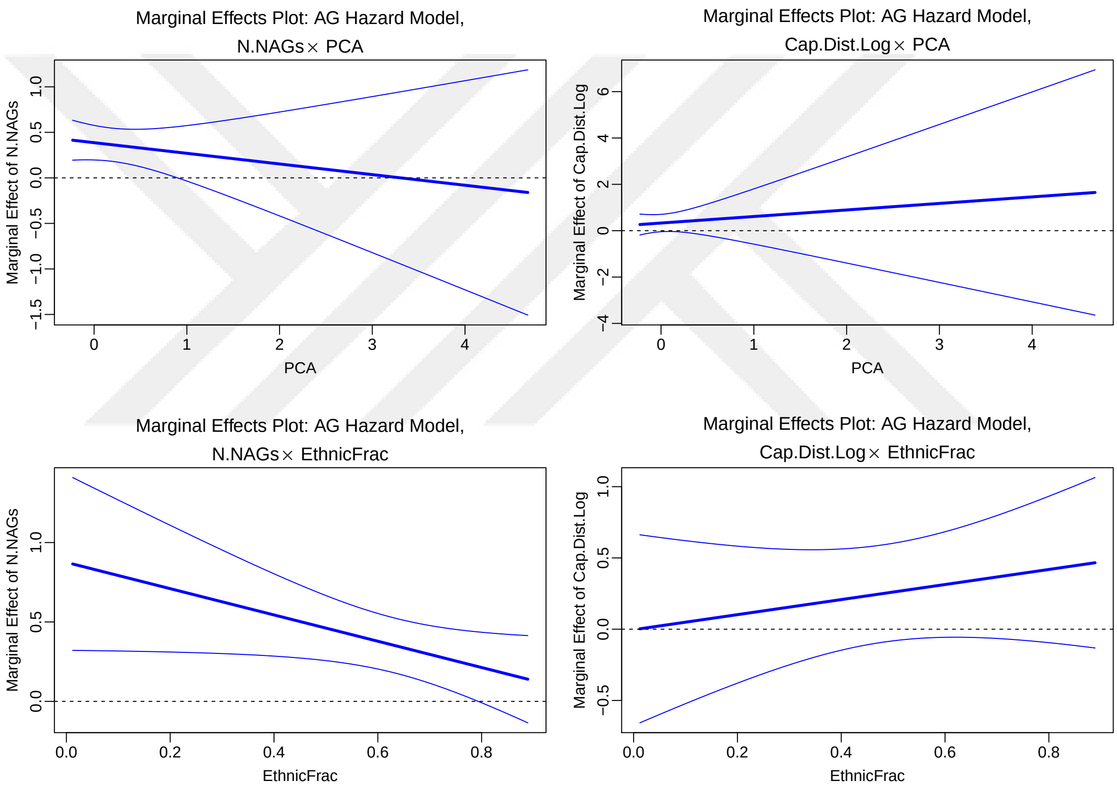

Figure 3.4: Marginal effects plots, AG hazard model.

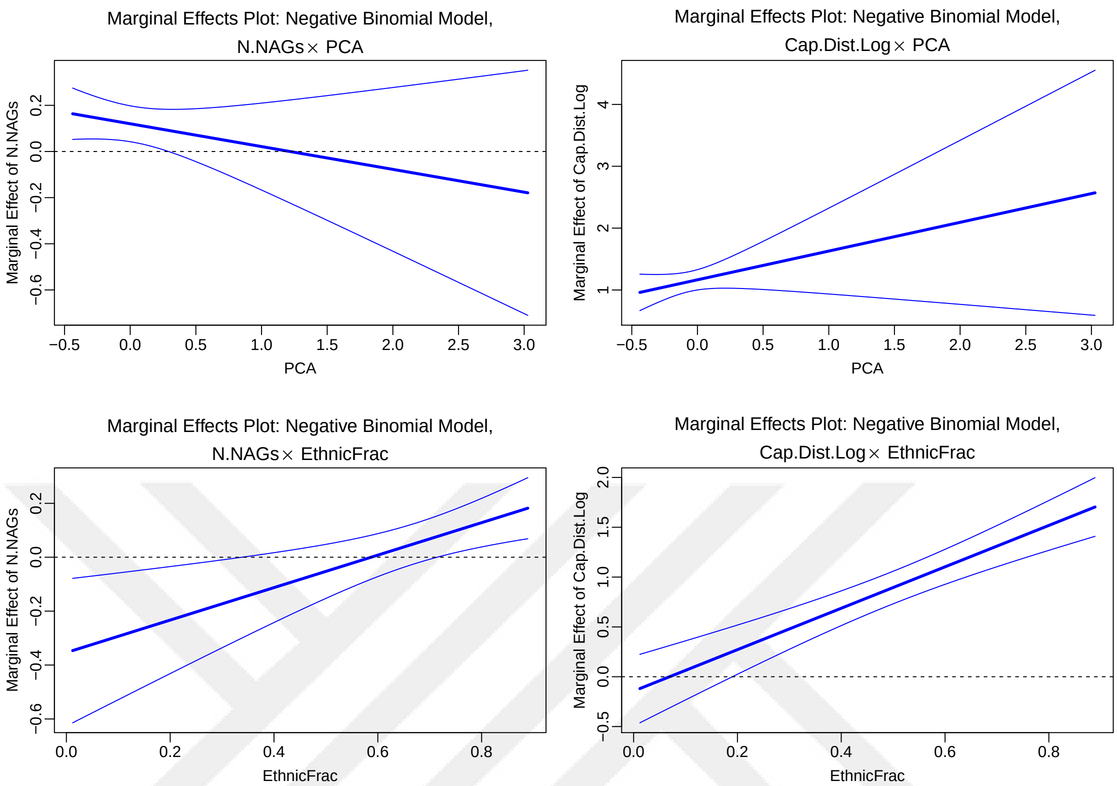

Figure 3.5: Marginal effects plots, negative binomial model.

is highly conditional on ethnic fractionalization, providing strong support for the intelligence gathering hypothesis. In combination with the results of the AG hazard model interactions, it seems clear the number of insurgents exhibits a much stronger conditional relationship with ethnic fractionalization, suggesting that states resort to PGMs in response to multiple insurgent groups since it is difficult to monitor numerous NAGs, not necessarily because more rebel groups pose a greater military threat to the state.

The effect of NAG peripherality conditional on both the PC and ethnic fractionalization is positive and significant at nearly all values in the negative binomial models. The combined results of the interactions suggests that *Cap.Dist.Log* lends support to both hypotheses. The results of the interactions in the negative binomial models can be seen in Figure 3.5. Table A3 in the appendix reports the regression results.

3.2.3 The Comparative Explanatory Power of the Interacted Models

While these results seem to suggest that *N.NAGs* supports the intelligence gathering hypothesis and *Cap.Dist.Log* may support both theories, to fully understand whether the interacted effect improves explanatory power, I test the out-of-sample predictive power of the interaction terms and compare them to models without the interaction. Just as in the previous section, to estimate the predictive accuracy of the models, I train each model on 80% of the dataset and then attempt to predict the outcome of the remaining 20%. I calculate the mean squared error to determine the relative accuracy of models with and without an interacted term. I include controls for regime type and GDP/capita in all models. To ensure that the results are robust to different samples from the data, I run 1000 iterations and report the average MSE of the different samples.

In short, the outcome of the predictive tests generally support the conclusions in the previous paragraphs. The results can be seen in Table 3.2. In the hazard models, both interacted terms fail to improve the predictive power of *N.NAGs*. Logged distance to the capital, however, seems to support the threat multiplier hypothesis since the interacted term with the PCA scores better than the interaction with *EthnicFrac*. The negative binomial models provided clearer results. The mean squared error of both *EthnicFrac* interaction models performed better than the non-interacted ones, while the interacted term with the PCA harmed predictive accuracy.

These results suggest that *N.NAGs* supports the intelligence gathering hypothesis, rather than the military threat theory. The negative slope of the marginal effect of $N.NAGs \times EthnicFrac$ in the hazard model complicates interpretation of this effect, however, since we would expect it to increase. I would argue that as ethnic fractionalization increases, it begins accounting for the same variation as *N.NAGs*, causing *N.NAGs*' coefficient to decrease. Further tests are necessary to clarify this conditional relationship. *Cap.Dist.Mean*, on the other hand, did not decisively support one hypothesis over the other. The contradictory results between the hazard and negative binomial models likely reflects that NAG peripherality contains traces of both hypothesized causes.

	Mean Squared Error	
Model Formula	AG Hazard	Negative Binomial
$PCA + N.NAGs$	1.000	5.645
$PCA + N.NAGs + Interaction$	1.000	5.663
$EthnicFrac + N.NAGs$	1.000	4.958
$EthnicFrac + N.NAGs + Interaction$	1.000	4.949
$PCA + Cap.Dist.Log$	1.050	4.413
$PCA + Cap.Dist.Log + Interaction$	1.042	4.608
$EthnicFrac + Cap.Dist.Log$	1.057	4.187
$EthnicFrac + Cap.Dist.Log + Interaction$	1.589	3.592

Table 3.2: Mean squared error for interaction models.

3.3 Building a Full Predictive Model

These previous analyses have compared the explanatory power of two competing yet not opposed hypotheses on PGM participation in civil conflict, but do both theories still provide useful information when modeling PGM mobilization? To test this, I run AG hazard and negative binomial models on every possible combination of all variables included thusfar in this analysis. Once again, I use out-of-sample prediction and mean squared error to evaluate the explanatory power of each combination. On account of limited computing power, I could only run 100 iterations since testing each sample involves running over four-thousand different combinations of variables. If the results heavily favor one set of variables over the other, it suggests that one hypothesis is truly the mechanism at work. If a combination of variables from both hypotheses best explains the outcome, however, it suggests that both theories account for variation in the outcome and provide theoretical leverage on PGM involvement in conflict.

The results from the hazard models indicate that a very parsimonious set of variables best predicts PGM mobilization: *EthnicFrac*, *N.NAGs*, and *Cap.Dist.Log*.

This provides very strong evidence that, in mobilizing new PGMs, states are more interested in gathering quality intelligence than in the fighting capabilities of paramilitaries.

In the negative binomial, however, both hypotheses found some support. The best combination of variables for predicting the number of active PGMs in a country is: *InterstateConf.*, *ForeignCommits*, *N.RebelSups*, *Reb.Strength*, *N.Rivals*, *EthnicFrac*, *Cap.Dist.Log*, and the PCA. This larger, mixed combination aligns more with my expectations. States are more likely to utilize multiple PGMs both in response to the need for information as well as for their military capabilities. Interestingly, the results of this test did not include the number of insurgent groups as a predictor. While further, more rigorous tests would be necessary to better understand this omission, states may not mobilize PGMs in direct response to an individual insurgent group, but rather employ PGMs based on other factors.

Ultimately, these hypotheses are not mutually exclusive. These theories are more likely *complementary* mechanisms rather than opposed ones. Nevertheless, these results suggest that the intelligence gathering hypothesis better explains the decision to mobilize PGMs than the force multiplier hypothesis. Military threat plays a role in multiple PGM participation, however.

This chapter has further clarified why PGMs mobilize and what factors lead states to use multiple PGMs. The results from Chapter 2 were confirmed by improving the operationalization of the principal component. Comparatively, *EthnicFrac* better explains PGM mobilization and involvement in civil conflict than the latent military threat facing the government. Additionally, this chapter examined two strong covariates of PGM participation in civil conflict—*N.NAGs* and *Cap.Dist.Log*— and found a strong conditional relationship between the number of insurgent groups and the need to gather intelligence. The conditional effect of peripheral insurgent groups remains ambiguous, however. The final predictive comparison between the theories provides further support for the idea that PGMs are mobilized more for their intelligence gathering abilities rather than for their fighting capabilities. Yet states may rely on multiple PGMs to augment the state's military capabilities. The theoretical conclusions and contributions of these tests are covered in the final chapter.

Chapter 4

THEORETICAL DISCUSSION AND CONCLUSION

The analyses in the previous two chapters have shed light on a critical aspect of pro-government militias in civil conflict. In this final chapter I will detail the contributions of this project to the growing literature on PGMs as well as its shortcomings. I will also discuss what this project reflects about the discipline and potential avenues for future scholarship.

4.1 *Contributions to Existing Literature*

4.1.1 Theoretical Contributions

First and foremost, this analysis is the first to quantitatively examine PGM mobilization and involvement in civil conflict. Other research has shed partial light on the critical question of PGM mobilization, guiding this project. However, by explicitly analyzing the contexts in which states utilize PGMs, my analysis clarifies why PGMs join a conflict and provides stronger theoretical expectations for PGMs' purpose and behavior.

This project also elucidates how PGMs fit into the myriad of actors involved in internal violence. As I argue in chapter two and three, PGMs are primarily intelligence gatherers. In the aggregate, states deploy PGMs when the state faces problems monitoring insurgencies, not when it faces greater military threat. This is a critical finding for the literature on paramilitaries and calls for a revision in how we theorize about pro-government militias. A large portion of existing studies focus on the violent aspects of PGMs at the expense of their other functions. The temptation to treat pro-government militias as "just another armed actor" will lead to improper expectations and misguided analyses.

Although some analogy to NAGs may be possible, using research on insurgent groups as a guide for studying PGMs will lead to poor theory. Imposing the same expectations for violence on NAGs and PGMs fails to accurately conceptualize the purpose of PGMs in civil conflict. I argue that the violent, anti-insurgent functions of PGMs are secondary to the task of relaying quality information about insurgents and their supporters to the state, but this function has received scant attention in the literature. Rather than focusing on PGM violence in general and treating peaceful PGMs as anomalies, we should understand PGMs as a peaceful group first, then study what makes them more prone to extreme violence. Some scholars have already researched what leads to PGM violence (Stanton 2015; Cohen and Nordås 2015; Barter 2013), but future research should emphasize the intelligence gathering purpose of PGMs and craft careful theory that balances paramilitaries' violent with non-violent functions. The results of this analysis provide a better picture of why PGMs join a civil conflict, thus generating better expectations for pro-government militia behavior.

Existing studies of PGMs and state sovereignty focus on how, by subcontracting violent functions to PGMs, states sacrifice their monopoly on violence in the territory. While some qualitative literature has argued that pro-government militias challenge a state's control over violence within its territory, I argue that these concerns are overstated. If we accept that PGMs are primarily intelligence gatherers rather than rogue violent actors, the government should not have problems disarming or controlling these groups. Existing research has shown that PGMs are more likely to target civilians and violate rights when the state already engages in these activities (Stanton 2015; Cohen and Nordås 2015). Where states fail to demobilize paramilitaries, it may reflect the states' *unwillingness* to disarm them rather than an inability to do so.

Perhaps localized violence by the PGM is simply a byproduct of empowering them against the local insurgent group. Or it may serve as a deterrent and improve the PGMs threat credibility when punishing rebel supporters. Still it may reflect local competition for control or retribution against a rival as suggested by Kalyvas (2006). Ultimately, however, I would argue that PGMs may actually improve state capacity by assisting the state against insurgents, increasing improving state penetration through better monitoring, and in some cases, acting as a government proxy in distributing public goods. Although a deeper analysis and theoretical exploration on PGMs and state sovereignty is beyond the scope of my analysis, future research can and should explore these grander theoretical questions on the delegation of violence.

The argument that PGMs are primarily intelligence gatherers comes with reservations, of course. The results from chapter three indicate that a mix of intelligence gathering and military threat variables explains the states decision to employ multiple PGMs. The predictive negative binomial models indicate that both *internal* and *external* threats play a role in the decision to employ mulitple insurgent groups. This was the initial expectation for the predictive tests since the two hypotheses aren't mutually exclusive, and the existing qualitative literature lends support to both theories. The results of the negative binomial model should be regarded with slight skepticism, however. The shortcomings of this modeling strategy will be addressed in more detail in the following section, but briefly, the negative binomial model includes strong distributional assumptions and poorly handles autocorrelation, casting doubt on its ability to accurately reflect the data generating process in this circumstance. Nevertheless, future research may seek to understand how both internal and external military threat impact a states' decision to use multiple PGMs.

This analysis has also found three very strong correlates of PGM mobilization and participation in conflict, namely ethnic fractionalization, the number of active insurgent groups, and the distance of insurgent groups from the state capital. These three variables shed some light on the understudied process of pro-government militias joining civil conflicts and confirm that PGM mobilization responds to conflict-specific factors. Furthermore, these correlations could reveal PGMs to be a missing causal link in other relationships in civil war scholarship.

For example, Collier's (2004) treatment of conflict duration showed that ethnic fractionalization lead to longer conflicts. This finding was criticized since there seems to be no clear mechanism that links fractionalization with longer civil wars. Montalvo (2007) built upon this work, arguing that ethnic polarization extends the duration of internal conflicts. The results of my analysis combined with Aliyev's (2020) expectations indicate that PGMs may be that causal link. Higher levels of ethnic fractionalization in a conflict setting creates barrier for the state to gather intelligence on insurgent groups. Thus, the state mobilizes one or more PGMs to make up for this intelligence need. The decision to use PGMs, however, can lead to heightened ethnic tensions or exacerbate distrust between the state and the NAG, ultimately obstructing a peace agreement and leading to a longer conflict. A similar dynamic may operate in situations where an insurgent group is located in the periphery of the state. Perhaps the distance itself explains some of the difficulty of combating the NAG, but PGMs and their attendant dynamics may play a part in extending the duration of peripheral conflicts.

While clearer tests and theory is necessary to prove these causal stories, this connection highlights the contributions of this project to both the study of pro-government militias as well as larger studies on civil wars. Studying actors in isolation has been common in the literature, but civil conflicts are obviously much more complicated. Despite being studied as a cause themselves, PGMs could also be part of a causal chain leading to certain outcomes in conflict. Only by understanding the links between the actors and how actors interdependently influence conflict dynamics can we foster a better understanding of internal conflict. This project represents a step toward integrating different research agendas into larger theories of civil conflict. Further speculation about future research will be discussed in more detail in the third section of this chapter.

4.1.2 Methodological Contributions

This project also illustrates the theoretical usefulness of creative modeling strategies and predictive testing. Traditionally, a basic Cox proportional hazard model is the method utilized in studies of actor onset in civil wars. However, as mentioned in Chapter 2, the Cox model cannot handle multiple failure events or time-varying predictors, making it ill-suited to reflect dynamic processes such as civil wars. Although not without some limitations, the Andersen-Gill hazard model is a great improvement, accepting dynamic covariates and handling multiple failure events while modeling the interdependence inherent in those mulitple failures. As we continue pushing the boundaries of the discipline and seeking to explain more complex phenomena, the modeling strategies we use must also become more complex. Conflict studies and the International Relations discipline as a whole must seek to better reflect the actual data generating process in our modeling decisions. This analysis highlights the value of such a strategy.

Furthermore, I heavily relied on out-of-sample predictive power to generate conclusions in my statistical analysis. While unorthodox in International Relations, predictive accuracy is objectively a better measure of the relationship between two variables as opposed to the arbitrary 95% confidence interval cut off. For example, in Chapter 3, distance to the capital barely failed to reach "statistical significance" but greatly improved predictive accuracy in both the AG hazard and negative binomial predictive tests. The goal of empirical research is to accurately explain and reflect observed phenomena. Out-of-sample prediction compares models on the basis of their ability to do exactly that. Rather than discarding variables because they fail to attain a certain p-value, predictive tests focus only on whether the variable can help explain an outcome and includes them in a model on that basis.

4.2 Shortcomings and Areas for Improvement

4.2.1 Alternative Theoretical Arguments

This project has demonstrated interesting and theoretically useful relationships, but some criticisms must be addressed. First, regarding my theory, some may criticize the conclusion that PGMs are primarily intelligence gatherers rather than violent actors. Just because the state does not use paramilitaries when facing greater military threat, PGMs could still be used to intimidate or target insurgency-supporting civilians or as local defense forces to solidify the state's control over that territory and repel rebel incursions. While these are plausible explanations and have some support in the qualitative literature, I would argue that the only reason the state would allow a PGM to operate in this manner is because of its intimate local knowledge. A locally-recruited PGM can distinguish insurgents and their supporters from the rest of the population and levy credible threats against those supporters (Lyall 2010). Without the PGM's intimate local knowledge, the state would not mobilize the PGM. The key to understanding PGM mobilization still lies in their ability to relay quality intelligence to the state. Of course, better data on PGM violence and detailed case studies of the day-to-day operations of PGMs would shed more light on the actual behavior of PGMs. But based on the results of this analysis, states clearly utilize PGMs when the state faces problems gathering intelligence, not necessarily when they face greater military threat.

Additionally, as Carey, Colaresi, and Mitchell (2015) suggest, governments may desire to repress minorities or dissenters and will use PGMs to avoid accountability for this violence. A similar mechanism may operate in ethnically fractionalized societies. A highly diverse population may be more difficult to govern, especially within a civil conflict. A state may employ a PGM to violently repress minority groups, reflecting this difficulty. Similar to the explanation in the previous paragraph, this counterargument questions whether PGMs are used for their violent capabilities against civilians rather than against rebels. Ultimately, testing this alternative theory is beyond the scope of this paper and outstrips available data. Definitively answering this question would require detailed information on PGM violence which, currently, does not exist. Future research should seek to determine whether, contrary to my argument, PGM violence toward civilians is a key to understanding PGM formation and participation in civil conflict. I would argue, once again, that local knowledge is still paramount for understanding PGM mobilization.

Others may criticize the use of ethnic fractionalization as the primary explanatory variable for the intelligence gathering hypothesis. Ethnic fractionalization has been criticized for numerous reasons, most commonly for being fixed through time and for lacking explicit causal mechanisms (Cederman and Girardin 2007). My implementation avoids the first problem through utilizing the *Historical* Index of Ethnic Fractionalization, as detailed in Chapter 2. The causal mechanism linking ethnic fractionalization and difficulty gathering intelligence could be improved, however.

The basic logic of this relationship states that group differences such as language, culture, or custom increases the cost of monitoring these groups. This effect should be even more pronounced in a conflict context where violence can exacerbate inter-group tensions. Still, the cost of gathering intelligence changes based on numerous other factors such as civilians' willingness to cooperate with the state, the nature of the differences between the state and minority group, how effectively the state has penetrated the target territory, among others. Although measures of some of these concepts are available, I was unable to model this more sensitively on account of data incongruencies, missingness, and the difficulty of matching these ethnic or language differences to each individual insurgent group then aggregating to the country-year level. Updated data on rebel strength and finer-grained data on both PGM and NAG location would enable future researchers to explicitly model these causal mechanisms, leading to stronger inferences and clearer conclusions.

Although ethnic fractionalization fails to explicitly capture some of these causal mechanisms, I argue that, given the available data, it retains usefulness as a variable. Few other quantitative measures with the same level of coverage could better capture the difficulty a state has monitoring insurgent movements and supporters. Despite its imprecision, ethnic fractionalization still accounts for much of the variation that a more sensitive set of variables would reflect. Ultimately, the findings of this project serve as a basis from which future research can test alternative theories and further refine our understanding of PGM participation in civil conflict.

4.2.2 Methodology

Just as with the theory, the methodology of this analysis might also be improved. First, even though the Andersen-Gill hazard model captures the data generating process much better than a traditional Cox proportional hazard model, it cannot account for multiple failure events in the same year. Thus in the 32 observations where mulitple PGMs formed in the same country-year, I had to discard data, potentially losing out on important variation that could tell us more about when states will mobilize multiple PGMs. I attempted to account for this with the negative binomial model, yet a more flexible hazard model may provide the best opportunity to explore multiple militia mobilization. Currently, I am not aware of a better model for this research question.

As mentioned in the previous chapter, the negative binomial model also exhibits some limitations that could be improved. First, the negative binomial model assumes independent observations, posing significant obstacles when used with time-series cross-sectional (TSCS) data. Although I sought to account for temporal autocorrelation with the cubic spline, better methods such as gaussian processes deal with this problem much more effectively (Rasmussen and Williams 2006; Carlson 2018). Second, the distributional assumption of the negative binomial model may be improper for my data. Many countries in my dataset never use PGMs, resulting in a high number of zeros in the outcome variable. On account of the novelty of this project, I was unable to account for the different data generating process between states which might use PGMs and those that never will.

More broadly, this analysis is subject to standard criticisms such as improper specification and omitted variable bias. In response, I would argue that given the complexity of the data generating process and the simplification necessary to test this relationship quantitatively, most currently available modeling strategies will include some level of misspecification. Considering the currently available methods, I sought to minimize these problems as much as possible. Entirely avoiding this problem, however, requires building a model specifically for this question.

Furthermore, given the relative lack of research on PGM mobilization and paramilitaries in general, more study is necessary to determine the correct, theoretically motivated control variables. My chosen controls—V-Dem's liberal democracy score and GDP per capita—have been shown to influence a state's propensity to utilize PGMs and are exogeneous to my explanatory variables. As research on pro-government militias improve, we will better understand which control variables we should use.

The methodological shortcomings of this project highlight the current weaknesses of methodology in International Relations. Overcoming these shortcomings requires careful consideration of the underlying data generating process and the best way to accurately model it. Where existing strategies fail, we should seek to popularize new models and approaches. This will not only strengthen the methodological rigor of the discipline but will also improve theory and inferences.

4.2.3 Broader Criticisms

Beyond problems specifically with this analysis, PGM studies also exhibit some general problems. To begin, studying paramilitaries is incredibly difficult and the pro-government militias database (PGMD) reflects this difficulty. Naturally, both states and PGMs conceal their relationship and operations. Open-source information about paramilitaries, their members, location of operation, and violent behaviors is scarce. The PGMD has done a fantastic job finding basic information about these groups, nevertheless, data problems severely restrict the types of questions that scholars can address. As mentioned numerous times, PGM mobilization occurs in a highly complex environment, but given the current state of data on PGMs, opportunities for quantitative analysis are limited.

Furthermore, the analytical definition of a pro-government militia also poses problems for building general theories. To illustrate, the PGMD codes each of these groups as PGMs: the Arakan Tigers in Serbia and the Interhamwe militia in Rwanda which were both instrumental in genocidal violence; the Peshmerga in Iraq which was formerly an insurgent group until the Kurds gained limited territorial autonomy; Gladio which was a NATO-funded paramilitary during the Cold War that targeted communists in Europe; the Revolutionary Guard in Iran which is separate from the official military but protects the regime from potential coups; and numerous, small civilian defense forces. The vast differences between these groups and their purposes creates major problems for developing any coherent theory about PGMs in general. This even applies to my own analysis. Obviously, the Arakan Tigers or the Revolutionary Guard are not primarily intelligence gathering groups, they are violent groups assisted by their local knowledge. In order to generate better theories and expectations for general PGM behavior, the discipline must develop better classification strategies to separate these groups. Some scholars have already built classification schemes for PGMs. Carey, who directs the PGMD, distinguishes groups based on their connection to the government (Carey, Mitchell, and Lowe 2012). Biberman argues that inexperienced "rookie" groups behave differently from "veteran" groups (Biberman 2018). Staniland lays out six different categories classifying non-state militias according to the ideological configuration of the state and the sub-state group (Staniland 2017). This project also suggests that the difference between offensive and intelligence gathering groups could prove useful for analysis. Future research could also code subcategories of PGMs such as death squads, civilian defense forces, or coup-proofers. Currently however, the scarcity of data on paramilitaries prevents quantitative researchers from utilizing these classification schemes in a meaningful, analytical way. Future research on PGMs must embrace and critically engage with these differences to continue forming incisive hypotheses and building theory.

4.3 Future Research

Finally, this project offers numerous opportunities for future research. Improving upon this analysis, I plan to build a specialized model to better capture the complexity of PGM mobilization. I will account for the difference between states that may use PGMs and those that never will with a "zero-inflation" step. This, on its own, may represent a theoretically useful finding since the current literature has not determined what makes states likely to mobilize PGMs beyond regime type. Then, I will try to extend the AG hazard model to include mulitple failure events in the same year to better capture multiple PGM mobilization. Both of these methodological additions will better reflect the actual data generating process and strengthen the quality of my conclusions.

I also plan to explore the question of PGMs and state sovereignty in a separate project. Numerous clear opportunities exist to test state building outcomes with PGM data. Building on this analysis, the future project would seek to understand how the intelligence provided by PGMs impacts state penetration and service provision. Perhaps the analysis would offer support for my previous argument that PGMs pose very little threat to state sovereignty. Conversely, the negative externalities associated with PGMs such as popular distrust of the government or longer lasting conflicts may prove to be so pernicious as to decrease the state's ability to consolidate its control over its territory.

Numerous other questions have also be raised throughout this project that could guide future research. Further in-depth qualitative study of the day-to-day operations and functions of pro-government militias would shed light on the balance between the intelligence gathering versus violent aspects of PGM behavior. This research could also consider the effectiveness of PGMs as an intelligence force. Existing literature suggests that within a conflict context local actors may use a connection to the state to settle local feuds and to fight for local power (Kalyvas 2006; Jentzsch 2014). If the state uses PGMs as intelligence gatherers, does this information significantly improve the state's ability to combat insurgencies? Does quality local knowledge reduce civilian victimization and lead to selective violence just against rebels? Preliminary studies have shown that states are more likely to decisively defeat rebels when the state uses PGMs (Aliyev 2019), but explicitly linking this to PGMs' intelligence gathering functions could further illuminate how paramilitaries influence conflict dynamics.

Future research could also consider the impact of other explanatory variables such as regime instability, internal divisions in the government, or a history of conflict. Regime instability (Eck 2015) and internal divisions (Mazzei 2009) have been connected to PGM usage in the qualitative literature. Former conflicts may create the networks necessary for PGM formation (Ahram 2011b) and make states more prone to use paramilitaries. Scholars may also consider how different types of conflicts influence PGM mobilization. All of these variables provide a clear opportunity for quantitative research. As the literature better understands the various factors that contribute to paramilitary mobilization, we can develop a stronger, holistic understanding of this critical actor in civil conflict.

There are also opportunities to connect paramilitary studies to other topics in civil conflict research. Specifically, integrating studies of private military companies (PMCs) with PGMs would greatly advance the study of internal conflict. Both research agendas have considered similar topics such as moral hazard, civilian victimization, and conflict dynamics. The explicit military goals of PMCs may neatly complement the intelligence gathering purpose of PGMs. If a state wants to improve its military capabilities, it may resort to PMCs, but rely on PGMs for local intelligence. A relationship such as this would build on the findings of this project and lead to a more integrated understanding of both of these actors individually as well as civil conflict in general. These larger, cross-actor studies are significantly more difficult but would improve existing theories of internal conflict.

4.4 Conclusion

PGMs are a critical yet understudied actor in civil wars. Their involvement in internal violence has serious consequences for conflict dynamics as well as larger questions in political science. Existing research has considered numerous aspects of PGM behavior and their impacts on conflict duration and resolution, but has not yet quantitatively considered what factors lead to PGM participation in internal conflict. Furthermore, extant scholarship has focused on PGM violence, at the expense of their other functions. In this way, paramilitary studies draws on comparisons to other violent actors to guide the research agenda.

This project has proposed a novel interpretation of the role of PGMs in civil conflict. Treating PGMs as "just another violent actor" risks seriously misrepresenting the actual purpose of PGM mobilization, leading to misguided research and misleading conclusions. Using latent military threat and the difficulty of gathering intelligence as competing hypotheses, my analysis has shown that the level of military threat does not improve our understanding of initial PGM mobilization. Instead, states are more likely to utilize paramilitaries when they cannot effectively gather intelligence on domestic NAGs. PGM's local knowledge is the key factor that leads states to delegate sovereign tasks to these non-state actors. Threat variables may yet provide clues as to when states will use multiple PGMs, however.

Beyond this theoretical contribution, this project has found numerous correlates of PGM participation in internal violence and illustrated the value of underutilized modeling strategies and tests. Future research can further clarify this relationship and continue to create better theories of PGMs behavior and how states combat insurgencies.

BIBLIOGRAPHY

Abbs, Luke, Govinda Clayton, and Andrew Thomson. 2020. "The Ties That Bind: Ethnicity, Pro-government Militia, and the Dynamics of Violence in Civil War". *Journal of Conflict Resolution* 64 (5): 903–932.

Ahram, Ariel I. 2011a. "Learning to Live with Militias: Toward a Critical Policy on State Frailty". *Journal of Intervention and Statebuilding* 5 (2): 175–192.

— . 2011b. *Proxy Warriors: The Rise and Fall of State-Sponsored Militias.* Stanford, California: Stanford Security Studies.

Akcinaroglu, Seden. 2012. "Rebel Interdependencies and Civil War Outcomes". *Journal of Conflict Resolution* 56 (5): 879–903.

Aliyev, Huseyn. 2019. "'No peace, no war' proponents? How pro-regime militias affect civil war termination and outcomes". *Cooperation and Conflict* 54 (1): 64–82.

— . 2020. "Pro-regime Militias and Civil War Duration". *Terrorism and Political Violence* 32 (3): 630–650.

Andersen, Louise, Bjorn Moller, and Finn Stepputat, eds. 2007. *Fragile States and Insecure People? Violence, Security, and Statehood in the Twenty-First Century.* New York, N.Y.: Palgrave Macmillan.

Arel-Bundock, Vincent. *World Development Indicators and other World Bank Data.* `https://vincentarelbundock.github.io/WDI/`.

Arjona, Ana M., and Stathis N. Kalyvas. 2008. *Rebelling against Rebellion: Comparing Insurgent and Counterinsurgent Recruitment.* Unpublished Manuscript. Yale.

Army, U.S. 2014. *Insurgencies and Countering Insurgencies 3-24.* Washington D.C.: Department of the Army.

Barter, Shane Joshua. 2013. "State Proxy or Security Dilemma? Understanding Anti-Rebel Militias in Civil War". *Asian Security* 9 (2): 75–92.

BBC. 2014. "Profile: Who are the Peshmerga?" *BBC.* 12 August, 2014. `https://www.bbc.com/news/world-middle-east-28738975`.

Belge, Ceren. 2016. "Civilian Victimization and the Politics of Information in the Kurdish Conflict in Turkey". *World Politics* 68 (2): 275–306.

Biberman, Yelena. 2018. "Self-Defense Militias, Death Squads, and State Outsourcing of Violence in India and Turkey". *Journal of Strategic Studies* 41 (5): 751–781.

Biberman, Yelena, and Megan Turnbull. 2018. "When Militias Provide Welfare: Lessons from Pakistan and Nigeria". *Political Science Quarterly* 133 (4): 695–727.

Blocq, Daniel S. 2012. "The Grassroots Nature of Counterinsurgent Tribal Militia Formation: The Case of the Fertit in Southern Sudan, 1985-1989." *Journal of Eastern African Studies* 8 (4): 710–724.

Böhmelt, Tobias, and Govinda Clayton. 2018. "Auxiliary Force Structure: Paramilitary Forces and Progovernment Militias". *Comparative Political Studies* 51 (2): 197–237.

Campbell, Bruce B. 2002. "Death Squads: Definition, Problems, and Historical Context". In *Death Squads in Global Perspective: Murder with Deniability*, ed. by Bruce B. Campbell and Arthur D. Brenner, 1–27. New York, N.Y.: Palgrave MacMillan.

Campbell, Bruce B., and Arthur D. Brenner, eds. 2002. *Death Squads in Global Perspective: Murder with Deniability.* New York, N.Y.: Palgrave MacMillan.

Carey, Sabine C., Michael P. Colaresi, and Neil J. Mitchell. 2015. "Governments, Informal Links to Militias, and Accountability". *Journal of Conflict Resolution* 59 (5): 850–876.

Carey, Sabine C., and Neil J. Mitchell. 2017. "Progovernment Militias". *Annual Review of Political Science* 20 (1): 127–147.

Carey, Sabine C., Neil J. Mitchell, and Will Lowe. 2012. "States, the Security Sector and the Monopoly of Violence: A New Database on Pro-government Militias". *Journal of Peace Research* 50 (2): 249–258.

Carlson, David. 2018. "Advanced Methods in Comparative Politics: Modeling Without Conditional Independence". PhD thesis, Washington University Saint Louis. `https://openscholarship.wustl.edu/art_sci_etds/1517/`.

Carter, David B., and Curtis S. Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data". *Political Analysis* 18 (3): 271–292.

Cederman, Lars-Erik, and Luc Girardin. 2007. "Beyond Fractionalization: Mapping Ethnicity onto Nationalist Insurgencies". *The American Political Science Review* 101 (1): 173–185.

Clapham, Christopher. 1996. *Africa and the International System: The Politics of State Survival.* Cambridge, UK: Cambridge University Press.

Clayton, Govinda, and Andrew Thompson. 2016. "Civilianizing Civil Conflict: Civilian Defense Militias and the Logic of Violence in Intrastate Conflict". *International Studies Quarterly* 60:499–510.

Clayton, Govinda, and Andrew Thomson. 2014. "The Enemy of My Enemy is My Friend: The Dynamics of Self-Defense Forces in Irregular War: The Case of the Sons of Iraq". *Studies in Conflict & Terrorism* 37 (11): 920–935.

Cohen, Dara Kay, and Ragnhild Nordås. 2015. "Do States Delegate Shameful Violence to Militias? Patterns of Sexual Violence in Recent Armed Conflicts". *The Journal of Conflict Resolution* 59 (5): 877–898.

Collier, Paul, and Anke Hoeffler. 2004. "Greed and Grievance in Civil War". *Oxford Economic Papers* 56:563–595.

Collier, Paul, Anke Hoeffler, and Mans Soderbom. 2004. "On the Duration of Civil War". *Journal of Peace Research* 41 (3): 253–273.

Coppedge, Michael, John Gerring, et al. 2021. *V-Dem Country-Year Dataset v11.* Varieties of Democracy (V-Dem) Project.

Cunningham, David E. 2006. “Veto Players and Civil War Duration”. *American Journal of Political Science* 50 (4): 875–892.

Cunningham, David E., Kristian Skrede Gleditsch, and Idean Salehyan. 2009. “It Takes Two: A Dyadic Analysis of Civil War Duration and Outcome”. *Journal of Conflict Resolution* 53 (4): 570–597.

Davis, Diane E. 2009. “Non-State Armed Actors, New Imagined Communities, and Shifting Patterns of Sovereignty and Insecurity in the Modern World”. *Contemporary Security Policy* 30 (2): 221–245.

De Bruin, Erica. 2018. “Preventing Coups d’état: How Counterbalancing Works”. *Journal of Conflict Resolution* 62 (7): 1433–1458.

Donahue, John D., and Richard Zeckhauser. 2011. *Collaborative Governance: Private Roles for Public Goals in Turbulent Times.* Princeton: Princeton University Press.

Drazanova, Lenka. 2019. *Historical Index of Ethnic Fractionalization Dataset (HIEF) V2.*

Eck, Kristine. 2015. “Repression by Proxy: How Military Purges and Insurgency Impact the Delegation of Coercion”. *The Journal of Conflict Resolution* 59 (5): 924–946.

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

Gade, Emily Kalah, et al. 2019. “Networks of Cooperation: Rebel Alliances in Fragmented Civil Wars”. *Journal of Conflict Resolution* 63 (9): 2071–2097.

Ganser, Daniele. 2005. *NATO’s Secret Armies: Operation Gladio and Terrorism in Western Europe.* London: Frank Cass.

Gberie, Lansana. 1997. “The May 25 Coup d’Etat in Sierra Leone: A Militariat Revolt?” *Africa Development* 22 (3/4): 149–170.

Gleditsch, Kristian Skrede. 2007. “Transnational Dimensions of Civil War”. *Journal of Peace Research* 44 (3): 293–309.

Gleditsch, Kristian Skrede, Idean Salehyan, and Kenneth Schultz. 2008. "Fighting at Home, Fighting Abroad: How Civil Wars Lead to International Disputes". *The Journal of Conflict Resolution* 52 (4): 479–506.

Harbom, Lotta, Erik Melander, and Peter Wallensteen. 2008. "Dyadic Dimensions of Armed Conflict, 1946–2007". *Journal of Peace Research* 45 (5): 698–710.

Humphreys, Macartan, and Jeremy M. Weinstein. 2008. "Who Fights? The Determinants of Participation in Civil War". *American Journal of Political Science* 52 (2): 436–455.

Jackson, Robert H. 1990. *Quasi-States: Sovereignty, International Relations, and the Third World*. Cambridge, UK: Cambridge University Press.

Jentzsch, Corinna. 2014. "Militias and the Dynamics of Civil War". PhD thesis.

Jentzsch, Corinna, Stathis N. Kalyvas, and Livia Isabella Schubiger. 2015. "Militias in Civil Wars". *Journal of Conflict Resolution* 59 (5): 755–769.

Kaldor, Mary. 2012. *New and Old Wars: Organized Violence in a Global Era*. 3rd. Cambridge, UK: Polity.

Kalyvas, Stathis N. 2008. "Ethnic Defection in Civil War". *Comparative Political Studies* 41 (8): 1043–1068.

— . 2006. *The Logic of Violence in Civil War*. Cambridge Studies in Comparative Politics. Cambridge: Cambridge University Press.

Lyall, Jason. 2010. "Are Coethnics More Effective Counterinsurgents? Evidence from the Second Chechen War". *The American Political Science Review* 104 (1): 1–20.

Lyall, Jason, and Isaiah Wilson. 2009. "Rage against the Machines: Explaining Outcomes in Counterinsurgency Wars". *International Organization* 63 (1): 67–106.

Maoz, Zeev, and Belgin San-Akca. 2012. "Rivalry and State Support of Non-State Armed Groups (NAGs), 1946–2011". *International Studies Quarterly* 56 (4): 720–734.

Mazzei, Julie. 2009. *Death Squads or Self-Defense Forces? : How Paramilitary Groups Emerge and Challenge Democracy in Latin America.* Vol. 1st ed. Chapel Hill: University of North Carolina Press.

Montalvo, Jose G., and Marta Reynal-Querol. 2007. *Ethnic Polarization and the Duration of Civil Wars.* Tech. rep. World Bank. `http://www.jstor.org/stable/resrep02492`.

Peic, Goran. 2014. "Civilian Defense Forces, State Capacity, and Government Victory in Counterinsurgency Wars". *Studies in Conflict & Terrorism* 37 (2): 162–184.

— . 2019. "Divide and Co-Opt: Private Agendas, Tribal Groups, and Militia Formation in Counterinsurgency Wars". *Studies in Conflict & Terrorism* 0 (0): 1–28.

Pereira, Anthony W. 2003. "Armed Forces, Coercive Monopolies, and Changing Processes of State Formation and Violence". In *Irregular Armed Forces and Their Role in Politics and State Formation.* Ed. by Diane E. Davis and Anthony W. Pereira, 387–406. Cambridge, UK: Cambridge University Press.

Rasmussen, Carl Edward, and Christopher K. I. Williams. 2006. *Gaussian Processes for Machine Learning.* Chap. 6. Cambridge, MA: MIT Press.

Salehyan, Idean. 2010. "The Delegation of War to Rebel Organizations". *The Journal of Conflict Resolution* 54 (3): 493–515.

Salehyan, Idean, Kristian Skrede Gleditsch, and David E. Cunningham. 2011. "Explaining External Support for Insurgent Groups". *International Organization* 65 (4): 709–744.

San-Akca, Belgin. 2016. *States in Disguise: Causes of State Support for Rebel Groups.* New York: Oxford University Press.

— . 2009. "Supporting Non-State Armed Groups: A Resort to Illegality?" *Journal of Strategic Studies* 32 (4): 589–613.

Schubiger, Livia. 2021. "State Violence and Wartime Civilian Agency: Evidence from Peru". *Journal of Politics, Forthcoming.* doi:`https://doi.org/10.1086/711720`.

Singer, J. David, Stuart Bremer, and John Stuckey. 1972. "Capability Distribution, Uncertainty, and Major Power War, 1820-1965". In *Peace, War, and Numbers*, ed. by Bruce Russett, 19–48. Beverly Hills: Sage.

Stacklies, W., et al. 2007. "pcaMethods – a Bioconductor package providing PCA methods for incomplete data". *Bioinformatics* 23:1164–1167.

Staniland, Paul. 2017. "Armed Politics and the Study of Intrastate Conflict". *Journal of Peace Research* 54 (4): 459–467.

— . 2012. "Between a Rock and a Hard Place: Insurgent Fratricide, Ethnic Defection, and the Rise of Pro-State Paramilitaries". *The Journal of Conflict Resolution* 56 (1): 16–40.

— . 2015. "Militias, Ideology, and the State". *The Journal of Conflict Resolution* 59 (5): 770–793.

Stanton, Jessica A. 2015. "Regulating Militias: Governments, Militias, and Civilian Targeting in Civil War". *The Journal of Conflict Resolution* 59 (5): 899–923.

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

Thompson, William, and David Dreyer. 2011. *Handbook of International Rivalries.* CQ Press.

Tierney, Dominic. 2005. "Irrelevant or Malevolent? UN Arms Embargoes in Civil Wars". *Review of International Studies* 31 (4): 645–664.

Tilly, Charles. 1992. *Coercion, Capital, and European States: AD 990–1992.* Cambridge, MA: Blackwell Publishers.

— . "War Making and State Making". In *The Formation of National States in Western Europe*, ed. by Charles Tilly, 1–89. Princeton, NJ: Princeton University Press.

Tollefsen, Andreas Forø, Havard Strand, and Buhaug Halvard. 2012. "PRIO-GRID: A Unified Spatial Data Structure". *Journal of Peace Research* 49 (2): 363–374.

Weber, Max. 1946. "Politics as a Vocation". In *From Max Weber: Essays in Sociology*, ed. and trans. by H.H. Gerth and C. Wright-Mills, 77–128. New York, N.Y.: Oxford University Press.

World Bank. *Gross Domestic Product per Capita.* The World Bank Group. Accessed October 15, 2020.

— . *Military Spending.* The World Bank Group. Accessed October 15, 2020.

APPENDIX

	AG Hazard		Negative Binomial	
	(1)	(2)	(3)	(4)
N.NAGs	0.328***		0.081***	
	(0.0795)		(0.030)	
Cap.Dist.Log		0.350*		1.045***
		(0.195)		(0.076)
V-Dem	0.254	−0.021	0.662***	0.781***
	(0.785)	(0.777)	(0.194)	(0.166)
GDP/capita	−0.129	−0.098	−0.180***	−0.162***
	(0.147)	(0.148)	(0.043)	(0.038)
Constant			−1.444***	−1.601***
			(0.113)	(0.113)
Observations	1842	1805	962	925
Log Likelihood	−505.64	−496.55	−1,599.144	−1,463.701
θ	0.830***	0.845***	2.463***	4.025***
	(0.277)	(0.296)	(0.226)	(0.452)

Note: The reported log-likelihood of the AG hazard models is penalized.

*p<0.1; **p<0.05; ***p<0.01

Table A1: Base models including the number of insurgent groups and distance to the capital.

	(1)	(2)	(3)	(4)
N.NAGs	0.387***	0.875***		
	(0.097)	(0.282)		
Cap.Dist.Log			0.330*	−0.004
			(0.191)	(0.343)
PCA	0.288		−1.865	
	(0.445)		(4.287)	
EthnicFrac		1.724**		−2.88
		(0.684)		(4.55)
V-Dem	0.498	0.452	0.207	0.088
	(0.801)	(0.793)	(0.808)	(0.786)
GDP/capita	−0.182	−0.140	−0.126	−0.072
	(0.149)	(0.150)	(0.151)	(0.153)
N.NAGs × *PCA*	−0.117			
	(0.151)			
N.NAGs × *EthnicFrac*		−0.827*		
		(0.424)		
Cap.Dist.Log × *PCA*			0.282	
			(0.575)	
Cap.Dist.Log × *EthnicFrac*				0.528
				(0.614)
Observations	1723	1723	1687	1687
Penalized Log-Likelihood	−482.171	−479.048	−472.75	−471.531
θ	0.769***	0.692***	0.780***	0.705***
	(0.268)	(0.253)	(0.285)	(0.272)

Note: *p<0.1; **p<0.05; ***p<0.01

Table A2: Interaction models, AG hazard.

	(1)	(2)	(3)	(4)
N.NAGs	0.120^{***}	-0.354^{**}		
	(0.040)	(0.139)		
Cap.Dist.Log			1.165^{***}	-0.144
			(0.084)	(0.179)
PCA	0.370		-2.241	
	(0.245)		(1.746)	
EthnicFrac		0.592^{*}		-10.600^{***}
		(0.337)		(1.733)
V-Dem	0.197	0.172	0.632^{***}	0.657^{***}
	(0.227)	(0.213)	(0.211)	(0.195)
GDP/capita	-0.072	0.048	-0.116^{***}	-0.082^{*}
	(0.049)	(0.047)	(0.044)	(0.043)
N.NAGs × *PCA*	-0.099			
	(0.089)			
N.NAGs × *EthnicFrac*		0.602^{***}		
		(0.199)		
Cap.Dist.Log × *PCA*			0.465	
			(0.327)	
Cap.Dist.Log × *EthnicFrac*				2.078^{***}
				(0.318)
Constant	-1.632^{***}	-1.755^{***}	-7.893^{***}	-1.246
	(0.136)	(0.251)	(0.488)	(0.944)
Observations	911	911	875	875
Log-Likelihood	$-1,487.569$	$-1,445.128$	$-1,353.519$	$-1,332.811$
θ	2.311^{***}	2.905^{***}	3.929^{***}	4.991^{***}
	(0.211)	(0.295)	(0.452)	(0.668)
Akaike Inf. Crit.	2,993.137	2,908.257	2,725.037	2,683.622
Note:				*p<0.1; **p<0.05; ***p<0.01

Table A3: Interaction models, negative binomial.