Violence, geography, and mobilization
A theory of insurgency

Author(s):
Schutte, Sebastian

Publication Date:
2013

Permanent Link:
https://doi.org/10.3929/ethz-a-010032050

Rights / License:
In Copyright - Non-Commercial Use Permitted
Violence, Geography, and Mobilization: A Theory of Insurgency

A dissertation submitted to

ETH ZURICH

for the degree of

Doctor of Sciences

presented by

SEBASTIAN SCHUTTE

M.A., ETH Zurich
M.A., University of Freiburg

Born August 21, 1979
Citizen of Germany

accepted on the recommendation of

Examiner: Lars-Erik CEDERMAN - ETH Zurich
Co-Examiner: Halvard BUHAUG - Peace Research Institute Oslo and NTNU
Co-Examiner: Ravi BHAVNANI - Graduate Institute Geneva

August 2013
In memory of my mother, Angelika Schutte
Acknowledgments

I owe a debt of gratitude to my supervisors, colleagues, friends, and family who inspired, taught, advised, and supported me during my PhD research. Most importantly, I would like to thank my advisor Lars-Erik Cederman for his support, patience, and guidance from the first to the last day of the project. The core theme of this study – changing loyalties in reaction to violence – is very much inspired by his works on emergent agency and grievance-based explanations for civil war onset. The data collection and relational database system created under his supervision (http://growup.ethz.ch) has also been an invaluable asset for this project. The other members of the PhD committee, Halvard Buhaug and Ravi Bhavnani, also laid the foundation of this study with their own research. Halvard Buhaug’s and Scott Gates’ 2002 publication on the role of geographic factors in civil conflict has sparked a rich and continuously progressing research program that combines quantitative modeling with geographic analysis. I have inherited both this selection of tools and a substantive interest in the spatial determinants of violence from him. Ravi Bhavnani’s simulation models of the micro-dynamics of civil conflict were also a great source of insight and our discussions of related research projects have been most inspiring and enjoyable. A former member of the research group, Nils Weidmann, also deserves a boatload of gratitude for his substantive advice, limitless patience, and continuous tech support. Hiring him was a very smart move by the University of Konstanz and I am sure that many generations of students will benefit from his teaching. Luc Girardin has also been very helpful with regard to all IT aspects of my research and especially in brushing up my Java skills. I would also like to thank Livia Schubiger for her
helpful feedback on early drafts and her suggestions for solving some of the methodological challenges.

Camber and Amy Warren were a major source of inspiration and fun during Camber’s Postdoc at the ICR group and while I’m happy they’ve found good jobs in the US, I regret that the pond is between us. My former office mate Christa Deiwiks helped with technical and motivational obstacles in the pursuit of this project. Thanks for all the jokes that made rainy days go by so much faster. Thanks to Karsten Donnay for his contribution to making our design for causal inference in conflict events data work, and for his collaborative spirit. Thanks for the time, brain power, and rocket science skills that you have invested! Julian Wucherpfennig and Philipp Hunziker let me tap into their deep knowledge of quantitative methodology whenever necessary. I was more than lucky – both figuratively and statistically speaking – to have not just one, but two probabilistic wizards willing to support ordinary mortals right next door. Thanks to both of you guys for sharing your insights on many occasions!

Seraina Rüegger, Nils Bormann, and Manuel Vogt were a great PhD cohort and having such a smart and nice bunch around made life and research a whole lot more enjoyable. Their collaborative and supportive spirit and their helpful feedback on early drafts were a great boost to my morale during otherwise nameless days and (regrettably) unforgettable nights at my desk. High five to you guys! We’re almost done! Luckily, the gap that our cohort will leave in the research group is being filled quickly with equally gifted and likable people: Andreas Schädel and Micha Germann.

I would also like to thank my family, in particular my parents for their emotional, intellectual, and financial support of my education. They watched me wander through the social and behavioral sciences for years and their confidence in my ability to succeed far exceeded my own. You are the best! Finally, I would like to thank Elin Arbin for her help, patience, and linguistic advice.
Abstract

Civil wars are the most frequent and most severe type of armed conflict since World War II. Most civil wars are fought as insurgencies in which at least one military actor is not recognized as belligerent and relies on irregular tactics. Despite the public awareness of the ongoing war in Afghanistan, as well as the uprisings in Libya and Syria during the Arab Spring of 2011, the processes that drive violence, mobilization, severity, and outcome of insurgencies remain disputed in conflict research and largely elusive to quantitative analysis. Traditionally divided research programs focusing either on the socioeconomic conditions that foster war onset or the dynamics of violence within conflicts have not yet produced an integrated picture of how insurgencies develop over time and how they end. Filling this gap, this thesis articulates, simulates, and tests a unified model of insurgency that can be used to predict the severity and termination type of such conflicts.

The research was guided by three central questions regarding the interplay between geography, violence, and mobilization which jointly affect the aggregate outcomes. The first question is: What are the determinants of the types of violence applied in civil wars? Varying levels of military control, initial motivations of rebel organizations, and competition over resources have been linked to the occurrence of indiscriminate violence, i.e. attacks that are not limited to enemy combatants and also affect civilian bystanders. Analyzing large samples of conflict events from 11 cases of insurgency, I am able to show that a simple distance-decay mechanism explains the types of violence used in civil wars surprisingly well: As the distance to their power centers increases, both insurgent and incumbent tend to apply more indiscriminate violence. Actors
are more likely to apply violence selectively close to their power centers.

The second question is how indiscriminate violence affects civilian loyalties and mobilization in civil wars. Diametrically opposite effects have been described in the literature. Deterrence-based reasoning suggests that higher levels of violence undermine support for the adversary. Alienation-based theories point to reactive mobilization, i.e. civilians supporting the adversary to take revenge against the perpetrator. Based on an in-depth study of the war in Afghanistan, I find that reactive collaboration with the adversary is the predominant consequence of indiscriminate violence.

The third question is how these micro-mechanisms of civil war scale to the macro-level. Outcome and severity have been explained through military tactics and external support to the rebels, but geographic aspects have been widely neglected. Since indiscriminate violence is applied predominantly far away from the actors' power centers and leads to reactive support for the adversary, population imbalances within countries have a strong effect on both the levels of casualties as well as the military outcome of insurgencies. Accounting for population imbalances strongly improves statistical predictions both in-sample as well as out-of-sample. The estimated effects correspond to the results of an agent-based model that simulates how the micro-dynamics of insurgencies scale to the macro-level.
Zusammenfassung

## Contents

1 Introduction 1

2 Existing literature and unresolved questions 11
   2.1 General overview ............................................. 11
   2.2 The focus on geography ...................................... 12
   2.3 The focus on motivation ..................................... 16

3 An integrated theory 23
   3.1 Geography and indiscriminate violence ..................... 23
      3.1.1 Technological aspects .................................. 24
      3.1.2 Tactical incentives ................................... 26
      3.1.3 Cognitive constraints ................................. 28
      3.1.4 Institutional aspects .................................. 30
      3.1.5 Summary: geography and indiscriminate violence .... 31
   3.2 Indiscriminate violence and reactive mobilization ....... 32
      3.2.1 Legitimacy ............................................... 33
      3.2.2 Counterinsurgency ..................................... 35
      3.2.3 Collective action ...................................... 37
      3.2.4 Grievances ............................................. 40
      3.2.5 Summary: consequences of indiscriminate violence .. 41
   3.3 Theoretical synthesis: causes and consequences in one model 41
   3.4 Empirical strategy ........................................... 45

4 Simulating insurgencies in an agent-based model 49
   4.1 Introduction .................................................... 49
   4.2 Existing literature .......................................... 50
   4.3 From theory to model ........................................ 51
   4.4 Detailed overview ............................................ 53
      4.4.1 Initial configuration .................................... 53
      4.4.2 The simulated Loss of Accuracy Gradient ............. 54
      4.4.3 Reactive mobilization .................................. 56
      4.4.4 Running the simulation ................................ 56
      4.4.5 A simple example ....................................... 58
   4.5 Simulated macro-level predictions .......................... 59
   4.6 Discussion .................................................... 65

5 The Loss of Accuracy Gradient 67
   5.1 Introduction .................................................... 67
   5.2 Existing literature .......................................... 68
5.3 A distance-decay model ........................................ 70
5.4 Data and case selection ....................................... 73
  5.4.1 Explanatory variables ..................................... 75
5.5 The war in Afghanistan ........................................ 77
  5.5.1 SIGACT data ............................................... 78
  5.5.2 Operationalizing indiscriminate violence in SIGACT .......... 79
  5.5.3 SIGACT results ........................................... 81
5.6 Civilian casualties in African insurgencies ..................... 85
  5.6.1 GED data .................................................. 87
  5.6.2 Operationalizing indiscriminate violence in GED .......... 89
  5.6.3 GED results ............................................... 89
5.7 Discussion .................................................... 91

6 Civilian reactions to indiscriminate violence ................. 97
  6.1 Introduction .................................................. 97
  6.2 Existing literature ........................................... 100
  6.3 Matched wake analysis ...................................... 107
    6.3.1 Sliding window design ................................ 108
    6.3.2 Statistical matching .................................. 111
    6.3.3 Estimation of causal effects ......................... 113
    6.3.4 Limitations of the approach ..................... 114
  6.4 Theoretical expectations .................................. 116
  6.5 Data .......................................................... 118
    6.5.1 Temporal dynamics ................................... 118
    6.5.2 Coverage .............................................. 118
    6.5.3 Eligible event categories ......................... 121
    6.5.4 Results ............................................... 123
  6.6 Discussion .................................................. 130

7 Territorial balance: geography, outcomes, and casualties .... 135
  7.1 Introduction .................................................. 135
  7.2 Existing literature ........................................... 137
  7.3 Theoretical expectations .................................. 139
  7.4 Empirical analysis ......................................... 140
    7.4.1 Measuring territorial balance .................... 141
    7.4.2 Analyzing outcomes .................................. 142
    7.4.3 Analyzing casualties .............................. 148
  7.5 Discussion .................................................. 154

8 Discussion, conclusion, and outlook ......................... 157

9 Supplementary information .................................... 165
  9.1 Supplementary information for chapter 4 .................. 165
  9.2 Supplementary information for chapter 5 .................. 169
  9.3 Supplementary information for chapter 6 .................. 171
  9.4 Supplementary information for chapter 7 .................. 184
1 Introduction

Young as the century may be, it has already witnessed the longest conflict in US history. Afghanistan, the battleground of the last Soviet war, has once more become the scene of a protracted irregular conflict following the US-led invasion. The 2003 war on Iraq – justified under the false claim that Saddam Hussein was in possession of weapons of mass destruction – has resulted in irregular resistance, ethnic violence, and a highly volatile security situation even after the withdrawal of American combat troops. Hundreds of thousands of civilians and combatants have lost their lives and many more their loved ones, their health, their property, and their foothold in civilian life. Beyond these theaters of war, irregular fighting in Chechnya and Ingushetia has long outlived the full-blown civil wars of the 1990s and continues to make headlines. Finally, the 2011 Arab Spring has seen a series of revolutionary uprisings in the Greater Middle East, two of which have devolved into lethal insurgencies in Syria and Libya with casualty estimates in the tens of thousands. In the wake of the Libyan civil war, a long-smouldering Tuareg insurgency has regained ground and momentum causing a French intervention in Mali.

All of these conflicts share of common type of warfare that has repeatedly attracted scholarly, military, and public attention: insurgency (Fall, 1965, McColl, 1969, Fearon and Laitin, 2003, Kalyvas, 2006, DOD, 2007). Broadly defined, insurgencies are civil wars in which the military combatants of at least one side are not recognized as belligerents and draw on irregular military tactics (CIA, 2009, 2). A less obvious feature of insurgency is its political dimension. In paraphrasing Clausewitz, Gahle (1964, 1) defines insurgency as implementing the policy of a party “by every means”, implying that political
activism and military action are pursued simultaneously by an ideologically committed party. In almost all cases, the aim of the struggle is to capture, control, and integrate the territory of a state into a new polity that appears more legitimate in the framework of the motivating ideology. Challenging colonial rule to implement political self-determination, replacing capitalism with socialism, reclaiming territory that has been invaded by a foreign country, and the establishment of Sharia law are political aims that have led to large-scale insurgencies. These “politicomilitary struggle[s]” over legitimacy (DOD, 2007, 1-2) usually take on the form of asymmetric conflicts: Rebel armies are often small at the beginning of the conflict and draw on guerrilla tactics to attack conventional military forces. In these early stages, the state is usually materially superior and employs a stronger conventional force to fight the rebels.

Far from being a new phenomenon, most civil conflicts in the post-WWII era were fought as insurgencies (Fearon and Laitin, 2003, Kalyvas, 2005, Fall, 1965, Kalyvas and Balcells, 2010). Insurgencies have created some of the most puzzling episodes of political violence in the twentieth century. On several occasions, the armies of major powers were beaten by irregular rebellions, for example in Algeria, Vietnam, and Afghanistan. Small numbers of combatants put up protracted fights against much larger regular armies despite heavy losses having been inflicted on them. In Algeria, 750,000 French soldiers fought 65,000 Algerian insurgents, inflicting gruesome losses in direct encounters, but also through torture and summary executions (Fall, 1965, Aussaresses, 2002, 48;120-121). Eventually, the 11-1 troop ratio did not lead to a stunning victory, but a defeat. On Cyprus, the British deployed 40,000 troops to combat 300 Greek insurgents and aimed for a political settlement five years later (Fall, 1965). In Vietnam, American and South Vietnamese forces certainly outnum-

---

1To illustrate the importance of insurgency, consider the following numbers: Fearon and Laitin (2003) introduced a dataset with 127 cases of civil war since 1945 that led to at least 1,000 battle deaths. Drawing on a subset of this data, Connable and Libicki (2010) identified 80 clear cases of insurgency. However, Kalyvas and Balcells (2010) report a declining importance of insurgency in the post-Cold War era in comparison to symmetric unconventional confrontations in civil wars. I am not following this distinction in this thesis, simply because the theory I am presenting is fully applicable to symmetric unconventional wars as well.
bered the irregular rebels and the North Vietnamese allies after 1965, but this did not prove decisive to the conflict (NARA, 2010). Nor did the fact that the US dropped more ordnance on Vietnam than in all theaters of World War II and in the Korean War combined (Clodfelter, 1995, Miguel and Roland, 2006). And yet, under superficially similar circumstances and in the same geographic region, British and American forces had beaten communist insurgencies in Malaya and the Philippines respectively. In Afghanistan, British and Soviet invasions were beaten by irregular rebellions, and the current US-led campaign is far from a comprehensive success.

These outcomes are not only puzzling in hindsight, but they also take decision makers by surprise. Some famous last words pay testimony to this: On a visit to South Vietnam in 1962, Secretary of Defense McNamara brushed off any doubts about the course of the war in Vietnam: “Every quantitative measurement we have shows that we’re winning this war” (Sheehan, 1988, 290). Eleven years and more than 50,000 American casualties later, the US finally withdrew from the country. On May 1, 2003, President George W. Bush announced on national television that “in the battle for Iraq, the US and its allies have prevailed” while standing in front of a star-spangled banner that read “Mission accomplished”. Weeks later, members of the disbanded Iraqi army took up arms to form the core of a protracted insurgency that gradually devolved into ethnic war, eventually claiming up to a million lives. Discussing the possibilities for crushing an uprising through the sudden application of massive force, military historian van Crefeld (2008, 244) used the 1982 Hama massacre in Syria as a prime example of successful deterrence. If it was not for the brutal ongoing conflict, one would call it an irony of history that van Crefeld picked Syria as an example: In response to an initially peaceful uprising in 2011, Bashar al-Assad responded with the use of heavy arms against civilians, as if following the scholarly advice. Instead of unconditional surrender, al-Assad ignited a full-blown insurgency that has so far resisted the government forces for almost two years with little outside support. Some 100,000 lives have been claimed by the fighting so far and there is no end in sight.
Unfortunately, there is little evidence that irregular civil wars are likely to cease any time soon, despite a decline in the overall levels of political violence in recent history (Harbom and Wallensteen, 2005). Moreover, in the last decade, strategic intervention by NATO members has become more common. In Libya, Mali, Pakistan, Yemen, and Somalia, air strikes against military forces and suspected members of terrorist organizations have been carried out repeatedly. Kilcullen (2010, 165) suggests that the current US counter-terrorism strategy fails to take into account a broader picture of interconnected Islamic insurgencies that are spreading internationally among groups that share cultural and linguistic ties. Calling for a more integrated understanding of insurgencies, Kilcullen (2010, 185) compares the historical implications of the uprisings in the Greater Middle East to those of the Soviet-backed Communist International in the first half of the twentieth century. But those seeking to overthrow their governments to established Sharia-based societies are not the only threats to peace. In a recent congressional hearing on the current legal status of the Authorization for Use of Military Force Against Terrorists (AUMF) that was put into effect shortly after the 9/11 attacks, the U.S. Assistant Secretary of Defense Michael A. Sheehan predicted that military action against al-Qaeda and its affiliates would go on for “at least 10 to 20 years”.

Similarly alarming is the fact that the original AUMF does not refer to specific military actors that can be engaged without congressional approval, but allows the US president to “use all necessary and appropriate force against those nations, organizations, or persons he determines planned, authorized, committed, or aided the terrorist attacks that occurred on September 11, 2001, or harbored such organizations or persons, in order to prevent any future acts of international terrorism against the United States by such nations, organizations or persons.” In combination with the recent advancements in using Unmanned Aerial Vehicles (UAVs) to

---


carry out precision strikes, this legal basis means that expanding the air strike campaigns in the Greater Middle East to new theaters of war has never been easier and cheaper for an American president.

In summary, both the recent history of irregular uprisings and the expectable future of continuous Western involvement in “out-of-area” wars will lead to large-scale human suffering and material destruction. With this expectation of protracted warfare looming, the question of how much reliable knowledge we have about the internal processes and drivers and the aggregate outcomes of insurgencies is becoming increasingly relevant. Put differently, before promoting armed intervention as a solution to contemporary crises, should we not ask what the outcomes of such interventions were and which trajectories these conflicts might take without such interventions? Clearly, a prerequisite for answering such questions is a theoretical model of how insurgencies develop over time, how new combatants are mobilized, and what aggregate outcomes can be expected under varying socio-economic and geographic conditions. Such knowledge is all the more needed as insurgencies display puzzling attributes with regard to their resilience to military pressure and technological advancements of the incumbent. In this regard, a better scientific understanding could provide valuable insights for preventing the worst disasters of the past from repeating themselves.

The questions of what drives civilians to take sides in civil wars and how rebels manage to sometimes win against overwhelming odds are certainly not new, but they are as relevant today as ever. Unlike previous generations of theorists and practitioners, the contemporary research program on civil wars is in a much stronger position to craft generalizable insights and models. The last decade has seen the rise of a cumulative and quantitative research program on the determinants and internal dynamics of civil wars and an unprecedented wealth of new and qualitatively superior data. Large-scale datasets holding information on all major conflicts since 1945 (Gleditsch et al., 2002), political and socioeconomic situations of politically relevant ethnic groups (Cederman et al., 2010), and the spatial configuration of subnational actors (Wucherpfen-
nig et al., 2011) have greatly contributed to a better understanding of the determinants of civil conflict. Large samples of conflict events from African civil wars have been coded with precise geographic coordinates and dates (Raleigh and Hegre, 2005, Sundberg et al., 2011) and opened a new route for quantitative inquiry. The leaks of hundreds of thousands of confidential military incident reports for Iraq and Afghanistan have provided the international research community with unprecedented insights into the spatial and temporal dynamics of irregular uprisings (SIGACT, 2010a,b).

Following this premise, this thesis develops and tests a theoretical model of the causes and consequences of indiscriminate violence in insurgencies. This theory helps to predict the approximate levels of casualties, and the military outcomes of insurgencies based on micro-level processes. Moreover, it has relevance for basic research in terms of helping to resolve three disputed questions on the dynamics of irregular civil war. First, what determines the quality of violence applied in civil conflict? Varying levels of military control (Kalyvas, 2006), initial motivations of rebel organizations (Weinstein, 2007), and competition over resources (Metelits, 2010) have been linked to the occurrence of indiscriminate violence, i.e. attacks that target both combatants and civilian bystanders, but a generally accepted theory is currently lacking. I use the term “selective violence” in this study to refer to attacks against individuals and groups in civil war that are conditional on past behavior or (para-)military affiliation. If individuals are harmed by military actors without being affiliated with the adversary and involved in the armed struggle, I refer to these attacks as “indiscriminate violence”.

Analyzing large samples of conflict events from 11 cases of insurgency, I am able to show that a simple distance-decay mechanism explains the types of violence used in civil wars surprisingly well: As the distance to their power centers increases, both insurgents and incumbents tend to apply more indiscriminate violence against both enemy combatants and innocent bystanders. Only close to their power center do actors apply violence selectively.

Second, how does indiscriminate violence affect civilian loyalties and mobi-
lization in civil wars? Diametrically opposite effects have been proposed by both scholars and decision makers. Deterrence-based reasoning suggests that higher levels of violence undermine support for the adversary (Lyall, 2009, Blackwill, 2010, van Crefeld, 2008, 241). Alienation-based theories point to reactive mobilization, i.e. civilians joining the ranks of the adversary to take revenge against the perpetrator (Ellsberg, 1970, Hastings, 2010, Kilcullen, 2009, 31-32). Based on an in-depth study of Afghanistan, I find that reactive collaboration with the adversary is the predominant consequence of indiscriminate violence in irregular wars for both incumbent and insurgent. This insight is especially important, as it rejects simplistic deterrence reasoning by underlining the importance of building legitimacy by refraining from the use of excessive force.

Third, how do these micro-mechanisms of civil war scale to the macro-level? Traditionally divided research communities have either focused on socioeconomic conditions on the country level that bring about civil conflict and shape its aggregate characteristics, or on the micro-dynamics of violence. Disaggregated computational studies (Cederman, 2008, Bhavnani et al., 2011, Weidmann and Salehyan, 2012) have recently paved the way for a more integrated understanding in that they simulate the complex pathways from micro-events to aggregate outcomes. I rely on a computational model and empirical tests of its predictions to show how the distance-decay of violence and reactive mobilization against the perpetrator affect the military outcomes and approximate levels of casualties in civil wars under varying geographic conditions. An econometric analysis of the effects of population imbalances corresponds to the simulated results and allows for statistical predictions of outcomes and casualties of insurgencies.

Taking previous contributions as a point of departure, this study therefore presents, models, and tests a new integrated theory of violence, mobilization, casualties, and military outcomes of insurgencies. It helps advance the young and rapidly developing area of quantitative research on the micro-dynamics of armed conflict by introducing a new method for studying causal relationships.
in conflict events data. Moreover, the study illustrates how agent-based models and probabilistic analyses can be combined to develop and test complex theories in social science.

The thesis proceeds as follows: Chapter 2 will review the existing work on insurgency and counterinsurgency. The effects of violence on civilian loyalties and the role of geography will be discussed in more detail, since these are recurring topics across various literatures. Most importantly, a series of unresolved questions will be identified which underline the need for a more thorough theoretical understanding. Chapter 3 introduces a theory of insurgency that integrates violence, systematic reactions to violence, military outcomes, and approximate levels of casualties. The theory draws on insights from counterinsurgency theory as well as theories of state reach and power projection. Empirical implications of the theory that deviate from competing explanations will be discussed in this chapter as well. Chapter 4 simulates the theory in an agent-based framework. Varying geographic conditions and different effects of violence on civilian loyalties are simulated to generate numerical predictions for competing theoretical explanations. The functional forms of these predictions serve as a frame of reference for the empirical studies. Chapter 5 is the first empirical chapter that tests one central implication of the theory in terms of the spatial distribution of different types violence in insurgencies. Again, several competing explanations and their implications will be considered alongside my own theory. In chapter 6, a new method is introduced for analyzing causal relationships in conflict events data and applied to the ongoing war in Afghanistan. In chapter 7, the implications of the modeled micro-dynamics for two macroscopic attributes of civil war – casualties and outcomes – will be tested. This chapter also demonstrates how the aggregated characteristics can be predicted out of sample based on exogenous, geographic conditions. Chapter 8 offers a discussion and conclusion for the study and points to possibilities for future research.

The following literature review provides a general overview of the existing research. Focal points of previous studies, such as the motivations for partici-
pating in insurgencies and the geographic conditions that make rebellion more feasible, will be discussed in more detail after that.
2 Existing literature and unresolved questions

2.1 General overview

By some accounts, irregular warfare by non-state actors is a distinctively new phenomenon of the post-Soviet era (Münkler, 2003, 7). But a closer look at the number of ongoing wars by type and year reveals that there is little new about these types of conflict from an empirical perspective (Harbom and Wallensteen, 2005). Theoretical attention to irregular war can already be found in the writings of Clausewitz (Freudenberg, 2008, 291) and the strategic reasoning of Napoleon Bonaparte (Tone, 1994, 3). The related literature reveals that insurgency and irregular warfare have received wide attention throughout the twentieth century, starting with works on the British-Arab guerrilla campaigns during Word War I. After World War II and the successful partisan activities against the Axis powers in Europe and south-east Asia, the topic received major attention in the communist literature which discussed tactical and ideological components: In Marx’s view, successful revolution requires a previous society to have developed advanced means of production that clash with the existing ownership structure, or more generally, the existing societal order (Marx, [1859] 1971, 9). While this prediction suggests that revolution in highly developed centers becomes more likely over the course of history, it also presents a bleak outlook for revolutionaries in less developed economies. The revolutionary period in Russia in the early twentieth century therefore provided an opportunity for a socialist takeover very much outside Marx’s
macro-historical reasoning, and the study of irregular warfare by communist theorists can be seen as a general attempt to craft revolutions out of limited military resources.

The successful uprising in China combined with the French defeat in Vietnam and the Cuban revolution led to a growing American involvement in south-east Asia and Latin America, and counterinsurgency doctrine moved into the focus of Western military advisors and scholars (Galula, 1964, McColl, 1969). During much of the Cold War, both the US and the Soviet Union invested material resources and strategic insights in irregular proxy wars, many of which began as insurgencies. While the classic international relations literature focused mainly on interstate war during much of the 1970s and 1980s, the end of the Cold War has moved civil conflict more into the focus of contemporary quantitative conflict research (see for example Fearon and Laitin, 2003, Buhaug and Rød, 2006, Kalyvas, 2006, Cederman et al., 2010). In the wake of the 9/11 wars in Iraq and Afghanistan, however, insurgencies in particular have once more gained public and scholarly attention (Kilcullen, 2009, Lyall and Wilson, 2009, Kilcullen, 2010, Connable and Libicki, 2010).

Clearly, all aspects of such a wide array of literature cannot be reviewed here, but there are two recurrent themes of central importance in all of these accounts. A careful review of these themes and a critical discussion of the limitations of the existing explanations will follow in the next two sections. After that, I will present a theory of the micro-processes in insurgencies and how they affect outcomes at the macro-level.

2.2 The focus on geography

The first recurrent theme is the focus on geography in the works of communist revolutionaries, counterinsurgency theorists, and contemporary conflict research. Lawrence’s strategic decisions already centered around the possibility of launching remote surprise attacks with irregular forces (Lawrence [1927] 1998, 157). Mao ([1938] 1967, 7) assumed guerrilla warfare to be most feasi-
ble when employed in large countries (such as China) where the conventional forces of the incumbent or the invader tend to overstretch their lines of supply. Guevara put a greater emphasis on escaping the state's reach through the utilization of difficult terrain (Guevara, 1961, 10).

Consequently, counterinsurgency theory and conflict research have also identified geography as a decisive factor. Military advisor Vann was troubled by the low degree of urbanization in South Vietnam in the early 1960s. With an estimated 85% of the population living in the countryside, guerrilla recruitment could take place largely unnoticed by the central government (Sheehan, 1988, 50). McColl (1969) offers a geographic model that identifies suitable territorial bases for guerrilla movements drawing on variables such as roughness of the terrain and remoteness. These variables have remained in the focus of basic research and are still being used in the quantitative analysis and modeling of civil conflict. Fearon and Laitin (2003) theorize that especially rough terrain allows a guerrilla movement to escape the state's reach, and Hegre and Sambanis (2006) confirmed that "mountainous terrain" is a variable robustly related with war onset.

With larger and more accurate events datasets becoming available to the research community, the question of where fighting takes place in armed conflict has emerged as an independent research topic. The spatial footprint of insurgency has been analyzed for Liberia (Johnston, 2008, Raleigh and Hegre, 2009) and Vietnam (Kalyvas and Kocher, 2009). Moreover, settlement patterns have been linked to both the likelihood of civil conflict as well as the occurrence of conflict events (Weidmann, 2009, Weidmann and Ward, 2010).

Geography also plays an important role in Kalyvas' (2006) theory of violence in civil wars. Drawing on a rationalist framework, Kalyvas (2006) proposes a model of how the application of violence in civil war varies spatially as a function of military control. In this framework, variation in military control causes violence to be applied either selectively against enemy combatants, informants, and collaborators, or indiscriminately against both combatants and neutral bystanders. A central prediction of this theory is the spatial separation of areas
where violence is applied selectively from those where violence is applied indiscriminately. Assuming that violence is used strategically, Kalyvas (2006, 202-207) suggests that highest levels of selective violence are applied in zones of predominant but incomplete control. Deviating from this pattern, indiscriminate violence is most likely to be applied in zones of little to no military control (Kalyvas, 2006, 223).

Kalyvas (2006) does not provide an independent model of how military control comes about, but acknowledges that it is largely endogenous to geography (Kalyvas, 2006, 132). At the same time, the theory makes the case that patterns of violence also follow the temporal variation of military control (Kalyvas, 2006, 119). While the theoretical value of this theory cannot be overstated, testing its observable implications across cases of civil war is extremely difficult. Even military actors in irregular wars struggle to paint a coherent picture of the geographic distribution of control. Based on an exceptionally rich dataset on levels of military control during the Vietnam War, Kalyvas and Kocher (2009) were able to test some empirical implications of Kalyvas’ (2006) theory quantitatively, such as the spatial separation of types of violence, but generally such fine-grained information is not available. Apart from data limitations, the notion that patterns of violence are rooted in levels of military control also implies a bleak outlook for attempts to predict areas that face an elevated risk of being affected by high levels of indiscriminate violence ahead of time. Therefore, the question of which exogenous geographic factors affect violence and overall outcomes of rebellion springs to mind.

The most widely accepted answer to this question points to the fact that distance and terrain limit the state’s potential reach. A classic model on the role of distance in interstate wars is presented by Boulding (1962) and is depicted in figure 2.1 on page 21. Boulding assumes that a state’s ability to project power toward an enemy is dependent on both its military strength and the distance that separates the adversaries. His notion of a “Loss of Strength Gradient” (LSG) assumes that for every unit of distance, a certain number of personnel must be subtracted from the fighting forces and added to the supply
troops. In theory, putting this relationship into numbers provides the basis for an exact calculation of the geographic limits in power projection. While heavily employed in the context of interstate conflict research (for example Lemke, 1995, Gilpin, 1981, 56-58), the model has also been modified to serve in the context of civil war research (Herbst, 2000, Buehaug and Gates, 2002, Herbst, 2004, Cederman, 2008, Buehaug et al., 2008, Buehaug, 2010).

By adopting Boulding's (1962) LSG model, Buehaug (2010) puts the notion of military control on an exogenous basis and shows that locations of major conflict zones in civil wars crucially depend on the government's capabilities: Strong states are more capable of repelling uprisings, forcing them to operate in the remote periphery while weak states also experience heavy fighting close to their power centers.

With a strong emphasis on geography, violence, and means of military projection, the existing literature has generated important insights. But while the significance of geographic remoteness for insurgent success has been widely acknowledged, the existing explanations fall short of painting a fully coherent picture:

First and most importantly, there is an unresolved puzzle of how state reach can explain outcomes of modern insurgencies: Boulding’s LSG provides an elegant and simple formalism that explains why rebellion is more likely and more successful in a state’s periphery. However, the two most prominent advocates of state reach explanations have stressed the mostly historical relevance of their research:

The airplane and now the missile have brought about a revolution of quite unprecedented dimensions [...]. For the air-born carrier or weapon, the world is an almost featureless globe: coasts, mountains, deserts, and forests hardly exist as long as there are landing strips on the other side of them. The intricate geographical structure of national power, therefore, which rests on the combination of sea-power with a very low LSG and land-power with a much higher LSG, has largely been swept away as far as air-power is concerned. Everywhere now is accessible to everybody; there are no nooks, corners, or retreats left, and no snugly protected centers of national power (Boulding, 1962, 272).
Similarly, Scott (2009, xii) remarks that his own theory of statehood’s inability to permanently integrate the south-east Asian highlands is largely irrelevant for the post-World War II period. Scott assumes that terrain friction accounts for a large variety of political attributes of the region, including the geographic shapes of states, the spatial distribution of literacy, and much of its early modern history of political violence. However, with the infrastructural and logistical advancements of the last sixty years, terrain friction has lost much of its significance, according to Scott.

Empirically we find that rebel victories have become more frequent over time in the past two centuries despite the overall trends toward more mechanized forces and increased long-range transport capabilities of conventional armies (Arreguin-Toft, 2001, Lyall and Wilson, 2009). Any air force today has the technical ability to project its force within the boundaries of its home state and successful insurgencies have been won even against superpowers with global strike capabilities, such as the United States and the Soviet Union in Vietnam and Afghanistan. The resulting puzzle is therefore: How is it possible that irregular insurgencies have become more successful despite substantial advancements in the power projection capabilities of conventional armies? I will propose a theoretical explanation for this phenomenon in chapter 3. In the next section, a second recurring theme in the literature will be discussed in more detail: the motivation of the combatants.

2.3 The focus on motivation

Beyond geography, every serious student and practitioner of insurgency has stressed its population-centric nature, or more specifically, the importance of generating public compliance and support. In his famous metaphor, Mao ([1937] 1961, 93) compared revolutionaries to the fishes swimming among the peasants as if they were the sea, implying seamless integration of the rebels with civilians. He also stressed that a large population was necessary for guerrilla war to succeed (Mao [1938] 1967, 11; 66). Guevara (1961, 2) echoes this
basic doctrine: "The guerrilla is supported by the peasant and worker masses of the region and of the whole territory in which it acts. Without these prerequisites, guerrilla warfare is not possible". The CIA's *Guide to the Analysis of Insurgency* identifies this aspect as a defining feature: "Insurgent activity – including guerrilla warfare, terrorism, and political mobilization [...] – is designed to weaken government control and legitimacy while increasing insurgent control and legitimacy" (CIA, 2009, 2). But how exactly can civilian support be generated by incumbent and insurgent? Solutions to this problem generally fall into two categories which speak, not surprisingly, to the timeless Machiavellian question of whether it is "better to be feared than loved" (Machiavelli, [1532] 1995, 51). Accounts opting for "fear" are plentiful.

According to some strands of literature in contemporary conflict research, individuals engage in rebellion if the prospects of economic revenue exceed the personal risks. Following this logic, Grossman (1999, 269) states that in many insurgencies the rebels are indistinguishable from bandits or pirates, and Collier (2000, 841) developed a model in which "rebellion is a special case of crime, with all the differences from conventional crime arising as a consequence of the particular loot sought by rebels". Clearly, the solution to stopping an insurgency within this model is simple: Increase the risk for the rebels through the increased application of military force. Certain leaders have accepted this mechanism. As Kissinger put it: "I refuse to believe that a third-class power such as Vietnam does not have a breaking point" (Greiner, 2009, 22). Greiner coined the term "tonnage ideology" referring to the assumption that a sufficient amount of attrition will translate into enemy surrender through a breakdown of either will or risk-reward calculations (Greiner, 2009, 23).

A formal framework that lends itself to modeling this trade-off is the classic collective action problem which has been repeatedly applied to insurgency (see Olsen, 1965, Tullock, 1971, Lichbach, 1995). Within this reasoning, increased risks should provide incentives for free-riding within the rebellion and therefore a weakening of its military capabilities. Translated into an observable implication, this approach suggests a positive association of the number of killed
insurgents with the probability of incumbent victory. The strong emphasis on the “body count” as a measurement of military success in the Vietnam War is a prime example of this logic in action (Greiner, 2009, 105).

A series of scholarly and military contributions strongly disagrees with this theory and its implications and, to paraphrase Machiavelli, opts for being “loved instead of feared”. Especially the counterinsurgency school has been focused on the problem of generating public support. The expression “winning hearts and minds” of the Vietnamese was made famous through its ritualized use by President Johnson. With the ongoing wars in Iraq and Afghanistan, the interest in this approach has reemerged and the emphasis on population-centric warfare is an essential part of today’s doctrine (DOD, 2007, 1-27). Military advisor John Paul Vann was one of the first outspoken critics of military tactics that alienated the civilian population in Vietnam and suggested that violence must be targeted selectively at enemy combatants to avoid civilian casualties (in Sheehan, 1988, 106; 317).

Along these lines, a considerable body of research argues that the use of indiscriminate violence has negative consequences for the perpetrator (Mason and Krane, 1989, Kalyvas and Kocher, 2007b, Wood, 2003, Kalyvas, 2006, 150). The killing of innocent civilians and the destruction of property can alienate the population from the attacker. Even if violence is targeted against combatants, killing them might help the opponent to recruit their kin. General Stanley McChrystal refers to this effect as “insurgent math” (in Hastings, 2010). Ellsberg (1970, 6) introduced an interesting metaphor. The rebels sometimes provoke military action by the state which in turn alienates the civilian population and aids the rebels. According to Ellsberg, this mechanism resembles the use of the opponent’s weight in Judo: instead of creating a comparative advantage for the attacker, the more weight they put behind their attack, the harder they fall themselves. Kilcullen (2009) coins the term “accidental guerrilla”, referring to elements of the local population that are drawn into the fight instead of being a priori adversaries of the incumbent. Clearly, this accidental process is closely linked to incumbent behavior in the field: As
civilians casualties mount and destruction of property continues, more locals might be willing to retaliate, independent of previous strategic loyalties. The *US Military Counterinsurgency Manual* also stresses the relevance of avoiding unnecessary destruction and violence and explicitly rejects the body count indicator (DOD, 2007, 5-27). While all of these accounts approach the problem from slightly different angles, the basic mechanism hereafter referred to as “reactive mobilization” is obvious: Instead of weakening the military opponent, violence can have the opposite effect. More civilians are repelled by the attacker and collaborate with the opponent.

Instead of being primarily interested in the effects of violence, Kalyvas (2006) provides elaborate insights into how violence is applied in civil wars, drawing on a stylized spatial model as discussed in the previous section. Kalyvas (2006) concludes that violence is used most selectively in zones of predominant but incomplete control. With regard to how violence feeds back into the dynamics of control, the theory makes the case for both coercive and alienating effects of indiscriminate violence. Kalyvas (2006, 144) assumes that indiscriminate violence is counterproductive for the perpetrator, as suggested by a wealth of qualitative evidence, but also points to the problems of testing such assumptions empirically in terms of the lack of adequate data and the impossibility of collecting reliable data under controlled conditions. Besides this mention of the possibility of reactive mobilization arising from the use of indiscriminate violence, Kalyvas (2006) ultimately assumes both violence and civilian collaboration to be endogenous to military control (pp. 12, 118-132).

Upon closer inspection, data quality is not the reason for Kalyvas’ neglecting reactive mobilization. Instead, the author conveys a causal explanation of violence being endogenous to control. Clearly, being in control militarily means having exclusive military presence in a given region. But if violence actually led to reactive mobilization, this effect would undermine military control simply by generating new combatants for the adversary. In this case, control would be endogenous to violence, as indirectly suggested by the counterinsurgency
Empirical studies on the effects of indiscriminate violence have produced contradictory evidence. A series of recent studies have found support for retaliation dynamics that drive civilian loyalties against perpetrators of indiscriminate violence (Condra and Shapiro, 2012, Braithwaite and Johnson, 2012, Linke et al., 2012), but the opposite effect has also been found (Lyall, 2009).

The systematic effects of indiscriminate violence therefore remain disputed, both from a theoretical and empirical point of view. Qualitative studies, military accounts, and anecdotal evidence suggest the existence of reactive mobilization as a mechanism in insurgency: Indiscriminate force creates a backlash in public reactions and leads to accelerated mobilization for the opponent. However, this logic is in direct contradiction to the assumption that increasing the risks for the opponent generates incentives for free-riding: Despite heavy losses and dim life expectancies for the individual combatant, some insurgencies have recruited at peak capacity:

The Loss Exchange Ratio in Vietnam is usually estimated to have been in the magnitude of 1:10 against the insurgents. Palestinian Intifada movements grew strong despite the smallest chance of victory. Afghan Mujaheddin are known for suicide tactics that completely deny worldly reward to the attacker. Counterinsurgency doctrine, long-range communication, and precision munitions seem not to have undermined the ability of insurgents to mobilize. Again, we find that the existing literature has developed the conceptual landscape considerably, but fails to answer another central question: How is it possible for insurgents to mobilize combatants despite heavy losses and dim chances of military success? The following chapter outlines an integrated theory connecting the geographic constellation of the military actors with the type of violence applied, as well as its consequence for mobilization.

\footnote{For a discussion of the causal pathway from control to collaboration, see Kalyvas (2006, 124).}
Figure 2.1: Boulding’s original LSG concept: States A and B have unequal military powers. Although A’s power declines over distance, it is still superior to B’s even in B’s capital. Boulding (1962, 232).
3 An integrated theory

The existing literature has identified geography and the motivation of combatants as decisive factors that explain both the onset as well as the outcomes of irregular civil wars. However, their exact roles remain disputed and two puzzles remain unsolved. An integrated understanding of insurgency must be able to explain (1) why increased state reach has not led to more successful counterinsurgencies in the past and (2) why rebel mobilization takes place even in the face of heavy losses and dim chances of success.

Using existing insights as a point of departure, I will attempt to provide such an integrated theory, arguing that the distance separating the power centers from the conflict zones affects the type of violence applied by the military actors, which in turn affects their ability to mobilize. This chapter will discuss this theory and its observable implications.

3.1 Geography and indiscriminate violence

While the general problems of balancing legitimacy and force in irregular war have been discussed extensively in various literatures, the geographic constraints on the application of conditional and selective violence have been widely neglected. Kalyvas (2006) points out that different types of violence in civil wars tend to occur at different locations, but attributes this effect to varying levels of military control which are endogenous to the conflict process. However, I argue that geography as an exogenous condition strongly constrains the application of selective violence for the military actors.

The most basic effect of geography on violence is that violence applied over
large distances or into unknown human and physical geographies tends to deteriorate in quality. While violence is generally highly selective and conditional close to the actors’ power centers, it becomes increasingly inaccurate as it is projected across space. A discussion of the technological, tactical, cognitive, and institutional mechanisms underlying this effect will follow.

### 3.1.1 Technological aspects

In a rarely remembered but perfectly practical definition of warfare, Boulding (1962, 266) reasons that war can be roughly defined as men throwing things at each other with malicious intent. Following this insight, Boulding traces the development of destructive capabilities in history in terms of projectiles being thrown at ever greater distances: “It starts with rocks, and it advances to spears and arrows, to cannons and rifles [...] and airplanes with bombs and guided missiles and nuclear warheads”. But this increase in range comes at a decisive price. From a technological point of view, violence follows a clear trade-off between range and accuracy. An intercontinental nuclear missile cannot be used to apply violence selectively. It is inherently indiscriminate. Similar in this regard but less devastating is conventional, strategic bombing: Explosives in free fall have a greater potential for indiscriminate destruction than direct small arms fire. Arendt (1970, 53) eloquently summarizes this effect by stating that the “inhumanity and destructive effectiveness [of weapons] increases in proportion to the distance separating the opponents”. Explicitly addressing the technological prerequisites for selective violence in counterinsurgencies, Vann famously remarked:

> This is a political war and it calls for discrimination in killing. The best weapon for killing would be a knife, but I’m afraid we can’t do it that way. The worst is an airplane. The next worst is artillery. Barring a knife, the best is a rifle – you know who you’re killing (in Sheehan, 1988, 317).
Clearly, Vann’s scale of selectiveness is inversely proportional to the range of the weapon systems he mentions. Despite the superficial differences between regular forces and rebels, these technological constraints are somewhat symmetrical for incumbent and insurgent:

At close range, both actors can selectively target individuals drawing on arrests and assassinations. At medium range, both actors risk civilian casualties in their attempts to kill enemy combatants using artillery strikes, tactical bombardments, and hit-and-run attacks. At long range, civilian casualties become inevitable in strategic bombing and international terrorism. Clearly, civilians can always be targeted on purpose at any range and by both actors, but the important insight here is that the ability to apply violence selectively becomes more and more constrained as the distance to the actors’ power centers increases. At greater distances, the only remaining options for the actors are indiscriminate attacks. Individual arrests and targeted assassinations are prototypical examples of selective violence, but their application is mainly confined to the areas under the actors’ control.

The advancement of arms technology has only superficially contributed to solving this trade-off. Today, manned and unmanned aerial vehicles are capable of hitting designated targets with great precision. But equating the ability to hit targets reliably with the ability to hit the right target effectively means confusing precision and accuracy. In measurement, accuracy refers to a system’s ability to reflect a true value, while precision refers to the system’s capability to reproducibly yield values with a constant offset to the true value. The dart game analogy can help clarify this distinction: Throwing darts accurately means to group them around the bullseye leading to a low average distance from the intended target. Throwing darts precisely means to tightly group them somewhere on the dartboard with similar distances from the intended target. With regard to the application of violence in (counter-)insurgencies, this distinction proves decisive: An accurate application of force would reliably affect actual enemy combatants and would be conditional on the severity of their actions while leaving innocent bystanders unharmed. A
precise application of force merely amounts to hitting what one is shooting at in combat situations. For example, calling in an air or artillery strike on a defined location in response to small arms fire will likely result in precisely hitting this location, but it might still harm bystanders. Clearly, technological advancements along the lines of guided munitions have enabled violence to be applied with greater precision, but not nearly as much with greater accuracy.

Following these considerations, we can assume that long-range attacks are – on average – less discriminate than short range attacks. Therefore, it is not necessarily the quantity of force that diminishes over distance as assumed by Boulding, but its quality: Arms branches with long range strike capabilities tend to apply force indiscriminately. In Vann’s words, a knife should be the weapon of choice in a political war, but a knife combines the highest accuracy with the shortest range. An airplane combines the lowest accuracy with the highest range. In between the two, there are rifles and artillery that follow the same trade-off.

3.1.2 Tactical incentives

Even if actors decide to not use aerial attacks or terrorist measures over long distances, they are presented with a tactical trade-off that also generates incentives for indiscriminate violence outside their areas of control.

Generally, combatants that advance into unknown and enemy-controlled territories find themselves exposed to an increased risk of ambush and attack. One way of mitigating this tactical disadvantage is the excessive use of firepower against suspected enemy positions. According to the motto of spending “shells, not men”, this approach became a standard procedure in Vietnam (Greiner, 2009, 38). Especially when terrain conditions made patrols on foot hazardous and time consuming, “harassment and interdiction fire” replaced close quarter engagement with enemy combatants (see Greiner, 2009, 150; Sheehan, 1988, 108; Lyall, 2009). Similarly, incoming sniper fire from civilian villages often led to troops calling in “close air support”, i.e. an airstrike against the entire village (Sheehan, 1988, 107). Clearly, random shellings or airstrikes in response
to small arms fire qualify as indiscriminate violence.\footnote{Lyall (2009) reports that this tactic was also used more recently in the Chechen wars.} These measures were not applied randomly all over the country, but in those areas where inhospitable terrain made rebel presence more likely and government presence insufficient to distinguish between rebels and bystanders.

Especially in situations were the identification of combatants is difficult, the trade-off between risking innocent casualties due to a false positive identification of enemy combatants and losing one’s life due to a false negative identification of an attacker generates incentives to “shoot first and ask questions later”. Generally, military actors put Rules of Engagement (ROEs) in place that allow for the use of lethal force against a perceived threat, even if the situation presents itself differently in hindsight. Reporting from the front lines of the 2003 US invasion of Iraq, Wright (2004, 53; 116; 139) describes the rapid escalation of the ROEs. Starting with strong restrictions on the use of lethal force, an initial set of ROEs only allowed US troops to shoot back when they were being shot at. After a series of lethal clashes with remaining Iraqi regular troops and irregular fighters, the ROEs were changed to declare anyone carrying a firearm a legitimate target. Finally, a third set of ROEs was passed that allowed the shooting of unarmed individuals wearing civilian clothing if they were suspected of serving as forward artillery spotters because they were talking into cell phones.

In areas of predominant state control, such as major cities, the pursuit of individual insurgents and their supporters is much less costly in terms of expectable incumbant casualties. Moreover, the use of firepower is also much more restricted in comparison to remote regions. Again, there is a striking symmetry between incumbant and insurgent with regard to deteriorating tactics as a function of distance: In rural environments where insurgents enjoy superior control, the identification and punishment of state collaborators can be performed on an individual basis. Outside these areas, insurgents lack knowledge of the local population and cannot rely on a dense network of informants. Direct attacks on incumbant forces are also extremely hazardous as
retreating into difficult terrain is problematic. As a result, insurgents usually attack the centers of state power less selectively. Terrorist tactics that allow for time-delayed explosions are a classic example of long-distance insurgent violence. Again, there is the same trade-off between range and accuracy.

3.1.3 Cognitive constraints

Beyond the technological constraints and tactical incentives, there is a cognitive dimension that makes indiscriminate violence more likely in areas unknown to the actors. The problem of limited information about civilian loyalties and the resulting incentives for indiscriminate violence have been discussed before, both from a psychological as well as rationalist point of view (Kalyvas, 2006, 69). The psychological mechanism at work, according to Kalyvas, is a fundamental distrust toward the civilian population, as well as frustration, uncertainty, fear, and anxiety resulting from the inability to tell apart foe and bystander (see also Greiner, 2009, 124). Moreover, insurgent tactics of surprise attacks and timely retreats make the clear identification of attackers extremely difficult. As a result, the local population is suspected of being actively involved with the uprising. This impression of meeting civilian resistance in pointless pursuit of an invisible enemy is commonplace in counterinsurgency campaigns: Ellsberg summarized the daily routine of US troops in the Mekong Delta as follows: “Foreign troops far from home, wearing helmets and uniforms and carrying heavy equipment, walking along dikes in formation and getting shot at every half hour mostly by ragged local irregulars firing from tree lines that bordered their homes” (Ellsberg, 2003, 167). Military advisor John Paul Vann described the problem similarly:

The Vietcong were so intermingled with the peasantry that the Saigon troops had difficulty distinguishing friend from foe. [...] How much more difficult it would be for Americans. The American soldier would soon start to see the entire rural population as the enemy [...]. ‘We’d end up shooting at everything — men, women, kids, and the buffaloes’ (in Sheehan, 1988, 383).

The term “Indian Country” Greiner (2009, 145) illustrates this general suspicion of US troops of the entire civilian population in the northern provinces of South Vietnam.
Especially in out-of-area operations, such as Afghanistan, fast rotation cycles add to the combatants' and commanders' inability to learn about local loyalties and heuristics for discrimination (see, for example Cowper-Coles, 2011, 167). By transferring commanders out of the theater of operation every six months, their ability to understand civilian loyalties is greatly reduced. In contrast, local resistance to occupying troops can continue to draw on local knowledge (Cowper-Coles, 2011, 62). Clearly, such a deployment schema greatly adds to the fundamental problem of fighting population-centric campaigns when knowledge of the population is missing. While acknowledging this problem, the solution proposed by the US Army Counterinsurgency Field Manual seems very ambitious for actors operating outside their cultural circle of origin: “Learn about the people, topography, economy, history, religion, and culture of the area of operations (AOs). Know every village, road, field, population group, tribal leader and ancient grievance. Become the expert on these topics” (DOD, 2007, Appendix A, Paragraph A2).

The French in Algeria, US troops in Vietnam, and the Soviets in Afghanistan found a variety of ethnic, linguistic, and cultural aspects that were mind-boggling to the outsider and hard to generalize from. Many historians and sociologists have struggled to paint a coherent picture of Afghan society and their most successful attempts require the introduction of social units and concepts that are foreign to most Westerners (see Dorronsoro, 2005). Mental access to foreign regions is difficult to acquire and communicate. Therefore, the cognitive ability to identify loyalties also diminishes as a function of distance. The constant threat of a looming surprise attack by largely invisible enemies can generate deep mistrust towards the entire civilian population and lead to trigger-happy reactions to suspected enemy presence. Language barriers further impede the ability to forge ties with the civilian population. Communication through interpreters is a poor substitute for the ability to understand connotations and subtext on a native level and to grasp a political atmosphere expressed in everyday life in direct communication with locals. In some instances, historical tensions between the local population and combatants can
also fuel prejudice, further complicating a cool-headed assessment of individual loyalties of civilians. Again, these effects are more likely to occur when cultural boundaries are crossed, i.e. at larger distances from the actors’ power centers.

3.1.4 Institutional aspects

One very counterintuitive aspect of the presented distance-decay of violence is that geography affects the behavior of combatants in ways that somehow circumvent top-down political and military decision making. While this effect should not be read as a geographic determinism, theater-level incentive structures and constraints in the applicability of accurate violence can travel up the chain of command and affect central decision making.

Troops in combat zones generally demand a free hand in dealing with threats to their lives. If operations take place in remote regions, the problem of limited military and political oversight greatly reduces the ability of higher echelons to micromanage activities on the ground. Greiner (2009, 175) describes this problem in the context of the “search and destroy” missions in the northern provinces of South Vietnam, where small groups of US personnel roamed the countryside for weeks on end without close central oversight.

As irregular war intensifies, the call for fewer restrictions on the use of force in an increasingly lethal struggle is usually passed up the chain of command. In the famous case of Raoul Salan, who served as commander of the counterinsurgency operation on Algeria, the insight that “with a partisan, one fights like a partisan” (Schmitt, 2007, 81) expresses the growing disregard for the rules of regular warfare. Salan followed this principle even against his own authorities and led a military insurrection in Algeria against the French administration. Later, he founded his own rebel organization and tried to regain power in Algeria. In a systematic study on civilian-military relations in wartime, Cohen (2002) refers to the “normal theory” – the idea that war might be started by a political decision maker, but it is then best left to the commanding general to fight and win it. Challenging this idea, Cohen (2002) engages in a series
of case studies, demonstrating the necessity for the continuous oversight of military actions by a civilian administration and emphatically argues against “leaving war to the generals”. With regard to Vietnam – the only irregular war under investigation in his study – Cohen traces the systematic escalation of the war back to the Joint Chiefs of Staff under President Johnson (pp. 175-178). Following the logic of bottom-up escalation, the Joint Chiefs forwarded the tactical necessities of deploying more forces to the theater of war to the political decision makers. Clearly, these examples illustrate that political decision making does not function in terms of a simple top-down process in counterinsurgencies and that armed forces can escalate conflicts by themselves.

3.1.5 Summary: geography and indiscriminate violence

I have discussed the distance-decay of violence in irregular civil wars and the underlying technological, tactical, cognitive, and institutional mechanisms. In summary, violence tends to deteriorate in quality when projected over large distances. Employing heavier arms from the power centers might inflict greater damage on the adversary, but they also cause greater levels of indiscriminate destruction. Granting a free hand to troops on the ground through liberal Rules of Engagement might reduce the number of combat losses, but it also generates incentives for preemptive and potentially indiscriminate violence, especially in regions where enemy presence is suspected. Limited cognitive access to foreign regions poses an additional problem for the conditional and selective application of violence. As combatants advance into unknown human and physical geography, simple heuristics for telling apart enemy and civilian generally replace the presumption of innocence. Top-down decision making can only balance these trade-offs instead of fundamentally resolving them and military considerations from the battlefield exercise pressure on the political leadership to develop strategies that inflict the greatest number of casualties on the enemy for the least number of losses.

These different mechanisms are equifinal in that they imply that violence becomes more indiscriminate in proportion to the distance separating the op-
ponents. Taking this effect into account, the advancements in power projection capabilities in the last century appear in a different light. Instead of military actors having extended their capabilities to control more and more remote territories, they have in fact extended their abilities to bring greater levels of destruction to more remote places. Explanations that solely focus on state reach, i.e. the quantity of violence that can be applied in civil wars, usually fail to consider this effect. The distance-decay model introduced by Boulding (1962, 232) lends itself to conceptualizing this effect. Instead of a “Loss of Strength Gradient” (LSG), I propose a “Loss of Accuracy Gradient” (LAG): Violence projected over longer distances deteriorates in quality and becomes less selective. In the next section, I will argue that the quality of violence has an important influence on mobilization in insurgencies.

3.2 Indiscriminate violence and reactive mobilization

In this section, I argue that the reaction to violence crucially depends on its quality, i.e. how selectively it is applied. Selective violence against combatants has a deterrent effect on other combatants and sympathizers in that it demonstrates the attacker’s capacity to target individuals conditionally on their behavior or affiliation. Indiscriminate violence is applied unconditionally and is therefore unjustified by any conceivable moral and political reasoning. Moreover, indiscriminate violence denies any payoffs even to those who comply with the demands of the attacker, as it is applied regardless of compliance. If this effect is taken into account, it becomes clear why insurgencies have succeeded even when the military odds were stacked heavily against them.

Theoretically, indiscriminate violence leading to increased resistance is a well-described, although sometimes disputed mechanism. Building on classic social contract theories, counterinsurgency theory, and recent conflict research, I argue that the predominant effect of indiscriminate violence is increased mobilization for the strategic adversary.
3.2.1 Legitimacy

Social contract theories are one of the central legacies of the Enlightenment and continue to inform contemporary discussions on the role of justice and authority in society (see, for example Rawls, 1971, Pinker, 2011). Social contract theories draw on central definitions and concepts, but widely disagree about substantive conclusions. For Thomas Hobbes and John Locke, the discussion of what constitutes a legitimate form of government starts with opposite descriptions of a “state of nature”: In Hobbes’ ([1651] 2009) view, such a state is marked by perpetual violence and stagnation in progress and development, as every individual is fighting for themselves and against all others. Consequently, personal safety, economic development, and political stability demand the strong hand of a central authority that keeps individual offenders and internal political challengers at bay. Hobbes’ Leviathan combines two axiomatic assumptions that are still at the heart of deterrence-based approaches to fighting insurgencies: a sovereign that is not constrained by law and a mechanistic idea of human that implies risk-averse behavior. In exercising authority, a strong sovereign is entitled to absolute authority as the sole source of the law. This impeccability extends even into the unjustified killing of citizens:

Nevertheless we are not to understand that by such liberty the sovereign power of life and death is either abolished or limited. For it has been already shown that nothing the sovereign representative can do to a subject, on what pretence soever, can properly be called injustice or injury; [...] And therefore it may and doth often happen in Commonwealths that a subject may be put to death by the command of the sovereign power, and yet neither do the other wrong (Hobbes [1651] 2009, 103).

Even in the most extreme cases of despotism, murder, and kidnapping, central authority cannot break the law.\footnote{Hobbes (2009 [1651], 103) continues to stress this point by reflecting on the killing of Uria by King David as described in the Old Testament (2 Samuel 11) and concludes that the killing might have been committed for morally questionable reasons. Therefore, King David might have committed a sin against God, but certainly not against the law, as changing the law on the fly was well within the King’s authority.}

From a Hobbesian point of view, the unlimited powers of the sovereign are unlikely to cause resistance among the ruled.
This expectation follows from the expectation that risk aversion and survival maximization drive human behavior as also commonly assumed in the modern rationalist framework. Consequently, Hobbes concludes that challenging the state by undermining its monopoly on violence can never be a legitimate endeavor. Thus, uprisings against the state are portrayed as indistinguishable from collective petty crime.

The Hobbesian stance, although strongly represented today in studies on the history of violence (Pinker, 2011, Gat, 2013), was never without opposition. An early critic was John Locke. Alarmed by the possibility of central authority abusing its power, Locke strongly emphasized the importance of legal boundaries for state power:

Where-ever law ends, tyranny begins, if the law be transgressed to another's harm; and whosoever in authority exceeds the power given him by the law, and makes use of the force he has under his command, to compass that upon the subject, which the law allows not, ceases in that to be a magistrate; and, acting without authority, may be opposed, as any other man, who by force invades the right of another (Locke, [1690] 1980, 103).

Moreover, Locke predicted that resistance to central rule was inevitable if governments engage in unlawful violence against their people:

But if either these illegal acts have extended to the majority of the people; or if the mischief and oppression has lighted only on some few, but in such cases, as the precedent, and consequences seem to threaten all; and they are persuaded in their consciences, that their laws, and with them their estates, liberties, and lives are in danger, and perhaps their religion too; how they will be hindered from resisting illegal force, used against them, I cannot tell. This is an inconvenience, I confess, that attends all governments whatsoever, when the governors have brought it to this pass, to be generally suspected of their people; the most dangerous state which they can possibly put themselves in (Locke, [1690] 1980, 106).

The early contributions to modern sociology preserved the idea of state authority relying on the legitimate use of violence. In Politics as a Vocation, Weber ([1919] 1992, 1) defines the state as a "human community that (successfully) claims the monopoly of the legitimate use of physical force within a
given territory. While Weber also points to the normative and personal traits of leaders that generate authority, their legality in terms of “rationally created rules” (Weber [1919] 1992, 2) remains a prerequisite for the the application of force.

In summary, two prominent lines of thinking rooted in the Enlightenment inform the current discussion on the driving forces in insurgencies. Building largely on an assumption of risk-averse human behavior and the role of state power to maintain order at all costs, the Hobbesian tradition stresses the importance of state strength in fighting internal challengers.

While some scholars continue to favor Hobbes’ depiction of a primordial state of nature over Locke’s (Pinker, 2011), the history of social contract theories has produced a clear verdict in condemning despotism. The more recent additions to this literature (for example Rawls, 1971) stress the importance of political and economic fairness in society. In this framework, freedom from arbitrary harm is among the most essential prerequisites for a societal order that can be called just (Rawls, 1971, 60). From this angle, the use of indiscriminate violence in irregular civil wars undermines the legitimacy of any perpetrator claiming a monopoly on violence. From a Lockean point of view, the predominant effect of indiscriminate violence is resistance, i.e. reactive mobilization on the part of the adversary.

3.2.2 Counterinsurgency

In addition to its importance for concerns of state legitimacy, military doctrines of the twentieth century have incorporated the effects of indiscriminate violence into their strategic calculus. An early expectation in military reasoning was that the increased lethality of modern arms would allow for rapid military victories. By targeting civilians indiscriminately, early military strategists reasoned, pressure would be exercised on the leadership of the adversary and enforce a quick surrender.

The prospects of causing large-scale destruction from the air looming in the late nineteenth century raised concerns which led to a banning of the “launch-
ing of projectiles and explosives from balloons, or by other new methods of a similar nature" at the first Hague Convention of 1899. Based on the experiences of World War I, aerial warfare was increasingly recognized as a future way to coerce entire states into surrender by inflicting unacceptable losses on their civilian populations. Douhet (1921) suggested that future wars would be decided mainly by the strategic bombing capacity of the adversaries. A breakdown of the civilian population’s will to sustain the war was the essential mechanism in his reasoning: Under intense bombardment the “time would soon come when, to put an end to horror and suffering, the people themselves, driven by the instinct of self-preservation, would rise up and demand an end to the war” (Douhet, 1921, 58). This basic logic continues to inform more recent scholarly work. Pape (1996, 16) summarizes the logic of coercion by means of aerial bombardment as the attempt to generate costs for continued resistance that exceed the costs of surrender. Based on his in-depth analysis of five cases of coercive air war, Pape concludes that the effects of bombardment vary widely, however: “Low to moderate levels of punishment inspire more anger than fear; heavy bombardment produces apathy, not rebellion [against the enemy government]” (Pape, 1996, 316).

Challenging the premise of air power theory, counterinsurgency doctrine (see Galula, 1964, Fall, 1965, Lansdale, 1972) developed rapidly during 1950s and 60s, questioning much of the conventional wisdom. Communist guerrilla activities in Malaya, the Philippines, and Vietnam underlined the importance of a combined political and military approach to containing insurgencies. The central premise of counterinsurgency theory is that the killing of innocent civilians and the destruction of property can alienate the population from the attacker and lead to increased support for the adversary. In a practical guide to counterinsurgency, Kilcullen (2010, 30) remarks: “But you have more combat power than you can or should use in most situations. Injudicious use of firepower creates blood feuds, homeless people, and societal disruption that fuels and perpetuates the insurgency”.

Variants of the expectation that indiscriminate violence leads to reactive
mobilization have been discussed throughout the literature. Kilcullen (2009) introduces a theory of rebel mobilization as a quasi-accidental process. While fighting a para-military opponent, incumbent forces cause civilian casualties and material destruction. This in turn fuels grievances that insurgents turn into support for their struggle. Others have identified reactive mobilization as an inevitable byproduct of fighting insurgents (McChrystal in Hastings, 2010): the family and kin of irregular combatants might seek revenge even if no bystanders were harmed. Attributing more intent to this mechanism, Ellsberg (1970, 6) suggests that rebels deliberately provoke incumbent forces to overreact in situations where negative externalities for the incumbent outweigh possible military gains.

Without discussing the exact causal pathway from violence to mobilization, modern counterinsurgency doctrine puts a strong emphasis on avoiding civilian casualties and unnecessary destruction (see DOD, 2007).

The ethical foundations of legitimate force in political philosophy and the utilitarian premise of counterinsurgency certainly differ, but their implications converge with regard to the application of violence in civil wars: Violence that is not applied selectively, i.e. regardless of affiliation or behavior, causes future resistance. As in the history of social contract theories, an initial plea for the use of unrestricted force was later challenged by more nuanced reasoning.

### 3.2.3 Collective action

The change in perspective that can be observed in political theory and military doctrine has also left its mark in conflict research. Classic rationalist reasoning informed the older scholarly works with respect to the motivation of individual combatants in rebellion. Following Olsen’s (1965) widely visible work on group behavior and collective action, a number of scholars have applied his framework to the special case of rebellion.

In summary, “collective action” refers to the problem that individual contributions to the provision of public goods are costly. Since public goods are provided to all, a better individual strategy for benefiting from them is to not
contribute to their provision and instead free ride on the efforts of others. But if free riding was the equilibrium strategy for everybody, then no one would invest in public goods provision. A solution to this problem is “selective incentives”, i.e. additional payoffs pursued privately by individuals engaged in public goods provision, which in this case is a side-product.

Tullock (1971) introduced a simple model for calculating the payoffs of participating in a rebellion against the state. Assuming that a successful revolution would increase public goods provision, Tullock (1971) concluded that this prospect was insufficient for motivating civilians to become rebels, since their individual participation was unlikely to substantially alter the prospects of successful revolution, while the risk of being killed in combat (i.e. the expected cost of participation) was high: “In sum, the theoretical arguments for the view that revolutions are carried out by people who hope for private gain and produce such public goods as they do produce as a byproduct seem to me very strong” (Tullock, 1971, 99). Such reasoning can also be used to predict the prospects of revolutions: “If most human beings would really rather be dead than red, then no society would be red. But in the real world most individuals care much more about their own welfare and survival than about public policy or the ideology of the society” (Olsen 1990 cited in Lichbach 1995, 16). Lichbach (1995) provides an extensive comparison between two prominent theories for explaining what determines individual participation in rebellion: the “Deprived Actor” versus the “Collective Action” perspectives. Arguing in favor of “Collective Action”, Lichbach (1995, 16-19) points to the fact that few nominal rebels actually participate in combat, interpreted as individual free riding. While rarely made explicit, this line of reasoning also issues a clear statement on how to fight rebellions. If the inclination of individuals to participate in rebellion is simply determined by how much the expected personal gains exceed the expected personal costs, then any rebellion can be stopped by inflicting prohibitive costs on the participants.

However, upon closer inspection, two objections can be raised against this conclusion. First, the collective action perspective suggests that rebellion can
only succeed by providing selective incentives to combatants, but it falsely
assumes that such incentives must be material in nature. However, the family
and friends of victims of violence might simply seek to avenge their innocent
loved ones. As Pinker (2011, 530) observes, "revenge is not confined to political
and tribal hotheads but is an easily pushed button in everyone’s brains". In
social and cognitive psychology, the desire for revenge has long been identified
as a strong motivator for violence (see McCullough et al., 2013) and has been
directly linked to terrorist mobilization (Speckhard and Alkmedova, 2006).

Arguably, indiscriminate violence, which lacks legitimacy and selectiveness
by definition, induces the desire to avenge the innocent. This mechanism has
profound effects on mobilization in insurgencies because it solves the collective
action problem by providing selective incentives to combatants: Those who
witness indiscriminate violence being applied become more motivated to fight
the perpetrator.

Second, the expected costs of fighting the perpetrator versus the costs of
compliance converge as violence becomes more random and indiscriminate.
This is due to the fact that neither compliance nor passivity offer protection
from the perpetrator when violence is applied regardless of (previous) behavior.
In this sense, indiscriminate violence imposes expected costs on both collabo-
rators and defectors, thereby undermining the logic of deterrence (see Mason
and Krane, 1989, Martinez and Morgan, 2011). The same effect can arise when
the costs of non-affiliation with any actor approaches the costs of affiliation
due to the lack of available protection (see Kalyvas and Kocher, 2007b).

In summary, the classic collective action problem has been applied to study-
ing insurgency and produced straightforward expectations: The higher the
costs of rebellion, the greater the imposed collective action problem and the
weaker the rebellion will become. But this reasoning neglects the notion that
revenge against the perpetrator of indiscriminate violence can serve as an in-
material selective incentive. Moreover, the implication that higher levels of
violence will lead to more effective deterrence also implicitly assumes condi-
tionality in the application of violence. In the case of indiscriminate violence,
such conditionality is absent by definition.

3.2.4 Grievances

As argued above, the implications of the collective action problem critically hinge on individual utilities: If revenge against the perpetrator is taken into account, reactive mobilization is the predominant reaction to indiscriminate violence. Moreover, Lichbach’s (1995) favoring of “collective action” over “deprived agency” as the premier paradigm to explain rebellion has been questioned. While a first generation of quantitative civil war studies heavily relied on material self-interest (Collier and Hoefler, 1998, 2004) and the feasibility of rebellion (Fearon and Laitin, 2003) as explanations for civil war onset, the underlying theoretical explanations were never without opposition. A key problem of such cross-national studies of war onset was that central variables were used as proxies for widely varying concepts (see Blattman and Miguel, 2010, 23). The reliance on over-aggregated and under-theorized indicators has been criticized in recent years (see Cederman and Girardin, 2007) and quantitative operationalizations of political and economic grievances have gained strong momentum (Cederman et al., 2010, 2011b, Wucherpfennig et al., 2011).

While research on the micro-dynamics of civil conflict has not yet fully embraced grievance-based explanations (see Wood, 2003), the question of how indiscriminate violence affects civilian loyalties has been explored quantitatively. A series of recent studies have found support for retaliation dynamics that drive civilian loyalties against perpetrators of indiscriminate violence. Analyzing instances of violence that caused civilian casualties in Iraq, Condra and Shapiro (2012) drew on precise geo-referenced data and found increased activity of the military adversary in response to indiscriminate violence. This finding was later confirmed by Braithwaite and Johnson (2012) who applied a non-parametric analysis of the spatio-temporal clustering of conflict events. Linke et al. (2012) focused explicitly on retaliation dynamics in Iraq and also found empirical support for reactive mobilization. The opposite effect, i.e. reduced activity of the adversary in reaction to indiscriminate violence, has also
been found in an in-depth analysis of the Second Chechen War (Lyall, 2009).

3.2.5 Summary: consequences of indiscriminate violence

Opposite effects of indiscriminate violence in civil wars have been proposed since the early days of the Enlightenment. One line of argument stretching from Hobbes’ *Leviathan* to modern air power theory to a materialist reading of collective action predicts that higher levels of violence will weaken the strategic adversary. This conclusion presumes that risk aversion and material self-interest are overriding motivators in human behavior.

Objections against this line of reasoning have been raised since the days of Locke. Questioning both the legitimacy as well as the effectiveness of brute force, modern social contract theories, counterinsurgency theory, and a nuanced reading of the collective action problem underline the importance of conditionality in the application of violence and predict reactive mobilization for the adversary in response to indiscriminate violence. Following this line of reasoning, I expect indiscriminate violence to have an alienating effect and lead to reactive mobilization. The next section will integrate the Loss of Accuracy gradient and reactive mobilization in a single theoretical model.

3.3 Theoretical synthesis: causes and consequences in one model

Two aspects of insurgencies have been discussed across various literatures: the geographic limits to power projection and the motivation of combatants. In the previous two sections, I have argued that the logistical limits to power projection are an insufficient explanation for why states fail to contain peripheral insurgencies. A series of cases where actors with global strike capabilities lost to guerrilla uprisings points to limitations of the paradigm. Moreover, while conventional power projection capabilities were massively extended during the twentieth century, insurgencies have become more, not less successful. Instead
of focusing on how the quantity of applicable violence diminishes as a function of distance from the power centers, I argue that the quality of violence declines as a function of distance due to a technological trade-off between range and accuracy, tactical incentives to apply violence preemptively under suspected enemy presence, and limited cognitive access to regions remote from the power centers of the actors. 4

Regarding civilians' reactions to different types of violence, two competing claims were reviewed. Siding with alienation-based theories, I have argued that reactive mobilization is the predominant effect of indiscriminate violence, since indiscriminate violence undermines the legitimacy of the perpetrator, generates selective incentives to avenge innocent victims, and fails to deter as it is applied unconditionally.

In two important respects, these assumptions deviate from Kalyvas' (2006) conclusions. First, although I agree that indiscriminate violence is most likely in areas of low military control, I arrive at this conclusion differently. Instead of assuming violence to be fully endogenous to control, I put a greater emphasis on the role of geography: Remoteness from the actors' power centers generally constrains the selective application of violence and can generate incentives for indiscriminate violence. Therefore, I do not fundamentally challenge Kalyvas' (2006) study, but attempt to put his typology of violence on an exogenous basis that is generalizable across conflicts. An empirical advantage of this approach is that it allows us to directly test the effects of exogenous geographic conditions, such as distances to the actors' power centers and terrain accessibility in the conflict zone on variation in violence. Second, while discussing the possibility of indiscriminate violence leading to reactive mobilization, Kalyvas (2006, 144-150) concludes that civilian loyalties are also generally endogenous to levels of military control. Investigating the causal pathways from military control to violence and civilian loyalties is certainly worthwhile in itself, but

---

4 Using geographic distance as a proxy for cognitive and linguistic barriers is not a new idea in itself. Alesina and Spolaore (2005) analyzed and modeled how nations of different sizes emerge in the international system. They also used geographic distance as a measure of declining cultural homogeneity (pp. 37-38).
attributing both violence and loyalties to military control falls short of providing a fully integrated picture of how outcome, severity, and mobilization can be explained from the initial conditions of irregular uprisings.

Following a Lockean reading of legitimacy, counterinsurgency theory, and recent empirical insights into civil war research, I assume reactive mobilization to be a central mechanism in civil war. I assume this mechanism to be so central that – in the long run – levels of military control are endogenous to reactive mobilization, which is a function of the type of violence applied by the actors. Applied violence, in turn, is strongly affected by geography.

I therefore propose a synthesis of Boulding’s distance-decay model and Kalyvas’ typology of violence: the loss of accuracy leading to reactive mobilization. Instead of the quantity of force that diminishes over distance as assumed by Boukling, I assume that its quality deteriorates. Moreover, I assume indiscriminate violence to create a backlash in popular opinions, leading to reactive support for the strategic adversary. The alienating effect of indiscriminate violence generally outweighs its deterrent effect, since even those who support the user of force can fall victim to indiscriminate attacks. This logic is illustrated in figure 3.1.

This theory provides a possible answer to the introduced puzzles: Instead of being deterred by the destructive capacity of the state and coerced into com-
pliance, civilians that witness indiscriminate violence might be more inclined to help the adversary. This could explain why certain insurgencies are capable of mobilizing despite high individual risks for combatants. Moreover, within this theory increased state reach does not necessarily imply more success in counterinsurgency operations. Mechanized armies and the high lethality of modern arms could generate strong reactive mobilization for the insurgent. The explanation would be in line with the historical trajectory of insurgencies having become more successful over the course of the twentieth century (see Arreguin-Toft, 2001, Lyall and Wilson, 2009). A number of empirical implications arise from this discussion. The first implication is the distance-decay of violence in insurgencies as expressed by hypothesis 1:

**H1:** The quality of violence in insurgencies declines as a function of distance from the attacker’s power center.

This expectation is tested on the level of single conflict events in 11 recent cases of insurgency in chapter 5. The second observable implication of the alienating effect of indiscriminate violence is expressed in hypothesis 2:

**H2:** Indiscriminate violence leads to more reactive collaboration with the adversary than selective violence.

Zooming in on the war in Afghanistan as a typical case of insurgency, chapter 6 will test this expectation on the level of single conflict events. Finally, I assume these effects to scale up to the macro-level and to affect military outcomes and levels of casualties in insurgencies. More specifically, I assume the outcome and severity of conflicts to be strongly influenced by the spatial distribution of population in war-torn countries. Hypotheses 3 and 4 communicate these expectations:

**H3:** The military actor that is closest to the bulk of the population is more likely to win the civil war.

**H4:** If opposing military actors are equally far away from the bulk of the population, the civil war will be most severe due to repeated cycles of reactive mobilization.
Figure 3.2: This figure illustrates the basic mechanism following the scheme of Coleman (1990). On the macro-level, the spatial distribution of the population within countries affects the military outcomes and overall severities of insurgencies. The causal pathway on the micro-level is also explained: Actors apply more indiscriminate violence far away from their power centers, which in turn leads to reactive mobilization. The associated hypotheses express the specified causal relations and will be tested empirically.

Hypotheses 3 and 4 are tested with a sample of 65 insurgencies and a quantitative indicator that captures population imbalances within countries in chapter 7. This combination of empirical investigations will form a coherent picture of how insurgencies develop, as depicted in figure 3.2. The links in the graph correspond to the individual hypotheses. Before testing the hypotheses empirically, I will introduce a simulation model in the next chapter to further illustrate the theory and generate numerical predictions for outcomes and casualties under varying geographic conditions.

3.4 Empirical strategy

Before diving into the details of the empirical analysis, I will quickly discuss the general strategy for testing the proposed theory. Clearly, testing for geographic variation in the type of applied violence and reactions to indiscriminate versus selective violence requires an integration of geographic information in the quantitative analysis. A recent turn toward the disaggregation of civil wars has provided much of the conceptual and technological groundwork. This type of research was pioneered by attempts to study subnational factors leading to civil war onset (Buhaug and Gates, 2002, Gates and Murshed, 2005, Cederman and Girardin, 2007, Buhaug et al., 2008) and research on the spatial determinants of conflict zones (Hegre et al., 2009, Raleigh and Hegre, 2009).
More recently, spatial data structures that code different geographic variables, such as distances to borders and the capital city, elevation, and population numbers, have been made available (Weidmann et al., 2010, Toellefse et al., 2012). Moreover, data on conflict events within civil wars have been made accessible to the research community that allow for a quantitative analysis of their spatial and temporal distributions (Raleigh and Hegre, 2005, SIGACT, 2010a,b, Sundberg et al., 2011, Leetrau and Schrod, 2013). Geographic information systems (GIS) allow us to calculate proximity and overlaps of such data.

The first two hypotheses of this study specifically ask under which conditions specific types of violent events are more likely to occur. In order to answer these questions, geographic information from multiple sources have to be combined. For data available in fixed raster cells, a ‘nearest-neighbor’ matching procedure was applied, i.e. a GIS operation that associates individual conflict events with geographic covariates, such as elevation and population numbers corresponding to their location. Distances from capital cities had to be calculated as well. In this way, GIS technology can be used to code context information for all instances of violence in any statistical sample, clearing the way for multivariate analysis.

Two general aims are pursued in the following empirical chapters. First, the causal relationships between geographic conditions and quality of violence as well as quality of violence and reaction to violence are analyzed. Both multivariate regression analysis (chapters 5 and 7) and statistical matching (chapter 6) are used for this purpose. But beyond causal analysis, this study also seeks to predict outcomes and casualties of insurgencies from their initial conditions. Therefore, the predictive performance of the employed statistical models is also assessed in chapter 7. A final and less conventional test of the presented theory builds on the output of the simulation model presented in the next chapter: By systematically varying simulated geographic conditions while keeping other factors constant, numerical predictions for casualties and outcomes of insurgencies are established. These predictions can be compared
to the predicted probabilities of statistical models. For this purpose, fitted models are used to predict casualties and outcomes based on artificial data. All explanatory variables are held constant at their means, while a quantitative indicator that expresses *territorial balance* is systematically varied. With regard to their functional forms, statistical and simulated predictions can then be compared. Corresponding functional forms indicate that the macro-level association of geographic conditions and outcomes as well as the micro-level simulation produce comparable results. This last test of correspondence between numerically simulated and empirically estimated predictions is necessary as it shows how the proposed micro-level interactions scale to the macro-level. The following chapter will introduce the simulation model and generate numerical expectations based on the presented theory.
4 Simulating insurgencies in an agent-based model

4.1 Introduction

This chapter presents a computational or “agent-based” model (ABM) that explains how the introduced mechanisms affect outcomes and casualties in insurgencies under varying geographic conditions. This model serves two basic purposes. First, it demonstrates the internal logic of the presented theory from chapter 3. Second, it generates numerical predictions that can be directly compared to the empirical results in chapter 7 (on applications of agent-based models, see also Van der Veen and Laitin, 2012, 284). However, the model itself is not critical to understanding the implications of this study and readers who prefer to interpret the empirical evidence directly can skip this chapter altogether. Readers who appreciate a computational demonstration will find that the stylized simulated predictions correspond well to the empirical findings presented in later chapters. More precisely, the model simulates reactive mobilization under varying Loss of Accuracy Gradients (LAG) and generates macro-level predictions for outcomes and casualties. Changes in the LAGs can be shown to account for variation in the severity and termination of simulated insurgencies. Correspondence with the empirical record is then established by comparing the observed frequencies of simulation outcomes to the predicted probabilities of statistical models. In this sense, the presented theory is put to a test which is commonly referred to as “generative sufficiency” (Epstein, 2007, 8). The word “test” in this context might be misleading as it differs
considerably from null hypothesis significance tests which are common tools in econometrics.

In a nutshell, “generative sufficiency” refers to a situation where a set of micro-mechanisms is capable of producing specific macro-level outcomes in computer simulations. Complementing probabilistic modeling and game theory, computational models have been used increasingly to explain macro-outcomes based on micro-interactions, starting with Schelling’s (1971) work on segregation. In more recent years, computational modeling has been applied to a variety of problems in social science, and in particular to the explanation of heavy-tailed empirical distributions, such as the distribution of incoming links on the World Wide Web (Barabási and Réka, 1999), as well as the frequency and severity of traffic jams (Nagel and Schreckenberg, 1992) and interstate wars (Cederman et al., 2011a). Such models have had an especially strong impact on the study of political violence and conflict research for several reasons. In general, simulation models allow researchers to capture typical civil conflict processes that are not easily accounted for in analytical models. For example, interaction under imperfect information, memory of previous interactions, and micro-level interactions leading to emergent behavior on the macro-level can be simulated. Finding equilibrium strategies in analytical models for such settings is a much greater challenge and often provably or practically impossible. These advantages make computational models valuable tools for the analysis of civil wars, i.e. for scenarios without clearly demarcated front lines, heterogeneous utilities, fluid loyalties, and restricted situation awareness for the actors.

4.2 Existing literature

Challenging prevailing theories that attribute rebellion to material incentives and military opportunities, Cederman (2008) introduced a model of insurgency

\(^1\)Of course, older strands in IR literature have also employed systemic theories and therefore built the conceptual foundations for some of the contemporary simulation models (see for example Waltz, 1959, Deutsch, 1969).
that focuses on the ability of culturally homogeneous peripheral groups to form coalitions. Bennett (2008) presents a simulation model for the early stages of insurgency that explicitly models the potentially coercive or alienating effects of state violence. Bhavnani et al. (2008) introduced an entire agent-based framework for studying the interactions between material incentives and cultural dynamics in bringing about civil conflict. Bhavnani et al. (2011) extend one central implication of Kalyvas' (2006) theory regarding the selective use of violence in areas of predominant but incomplete control to a three-actor setting and show that simulation results of the spatial distribution of violence are congruent with empirical findings. Drawing on a simulation model and geo-coded empirical data, Weidmann and Salehyan (2012) show that violence-induced migration in the war in Iraq accounts for the emergence of ethnically homogeneous neighborhoods in Baghdad. Utilizing geo-coded data on violent events in Jerusalem, Bhavnani et al. (2013) calibrated a simulation model with empirical data and then calculated counterfactual scenarios for different levels of segregation in the city. Beyond basic research, Kilcullen (2010, 192) calls for the use of simulation models for understanding insurgency. After laying out a comprehensive theory on counterinsurgency, Kilcullen (2010, 205) calls for a corresponding “complex adaptive systems” (CAS) (see Miller and Page, 2007) model that incorporates elements such as geography, civilian population, military actors, and limited information.

Since my theory is constructed around geographic variation in the accuracy of applied violence and changing loyalties resulting from experienced violence rather than anticipated payoffs, agent-based modeling was identified as a suitable methodology.

4.3 From theory to model

Following the theoretical discussion, the model assumes soldiers and rebels to fight each other while also competing for civilian loyalties. Insurgent forces, incumbent forces, and civilians are represented as individual agents. Each
civilians is associated with a value for government legitimacy. This value determines whether the civilians remain neutral or join one of the military actors. In every time step, each military actor gets a chance to use lethal force against another agent. This decision is based on the fraction of enemy combatants in the agent’s neighborhood and the local accuracy threshold. This value defines which fraction of adversaries in their neighborhood is sufficient for triggering an attack. This means that an agent in location X might only engage another agent in its neighborhood if there is a 90% chance of actually hitting the adversary and not a civilian. The same agent in location Y might nevertheless engage if there is only a 50% chance of exclusively harming the enemy. More specifically, the decline in accuracy for the military actors, i.e. the probability of engaging the adversary while sparing civilians, is expressed in two intersecting Loss of Accuracy Gradients. The slope of these gradients is defined by only one global constant: the territorial balance (TB). If the territorial balance favors insurgents, their accuracy declines less as a function of distance from the upper right corner of the simulated space, which means that they can apply violence more selectively within the simulated territory. Should the balance favor the incumbent forces, however, the insurgents’ accuracy declines rapidly, which leads insurgents to apply violence more indiscriminately. This aspect of the model captures a central element of the theory: the expectation that the ability to selectively engage the adversary while sparing civilians is a function of distance from the attacker’s power center.

In every time-step, the military actors have a chance to attack. If they choose to do so, a randomly chosen agent from their neighborhood is killed and is removed from the simulation. If this agent happens to be a civilian, all other civilians in its neighborhood are alienated from the user of force. Their value for government legitimacy is changed toward the military opponent of the user of force. This aspect of the model captures the second central element of the theory: the expectation that civilians witnessing indiscriminate violence are inclined to assist the adversary. Of course, this expectation is based on a Lockean reading of legitimate violence and a corresponding Hobbesian scenario.
without reactive mobilization can be simulated as well, which will be discussed below in more detail. While the simulation is running, all agents conduct a random walk across the simulated space to allow for repeated interactions between the actors under varying conditions. The simulation stops as soon there are no military actors of one side left. A closer look at the implementation of the model will follow in the next section.

4.4 Detailed overview

In order to simulate the introduced mechanisms, several parameters must be set for the the model to run. In this section, I will discuss these parameters. After that, a prototypical simulation scenario will be used to further illustrate the functionality of the running simulation.

4.4.1 Initial configuration

In order to initialize a simulation run, a fixed number of agents needs to be generated and distributed across the simulated space. In this case, “space” refers to a 50-by-50 cell grid, where each cell can hold one agent at a time. The total number of agents needs to be specified ahead of time. These agents are randomly assigned to locations in the simulated space according to draws from a uniform distribution. Their internal loyalties are drawn from a normal distribution with a given mean value for government loyalty and a given standard deviation for government loyalty. Vastly different populations of agents can be generated based on these parameters. For example, one can set the mean of government loyalty to 0.5 and the standard deviation to 0. This would give rise to a perfectly neutral population of civilian agents and no military actors. A higher standard deviation would lead to more and more military actors on both sides. Clearly, changing the mean value also enables us to create populations that are more biased towards either side.

Another parameter that is necessary to run the simulation is vision, which controls the size of the surrounding space that an agent can see. If agents have
a vision of 0, they cannot see anything that is happening to their neighbors. In such a case, military agents could not engage any opponents because they would not interact with any of their neighbors. More relevant settings allow agents to see what is happening in their local environment, but not everything that is happening in the entire space.

4.4.2 The simulated Loss of Accuracy Gradient

Of course, a crucial parameter in this simulated setup is the actors’ accuracy in the application of force. As argued in chapter 3, the accuracy of violence is largely dependent on the distance of the actors from their power centers. How is this mechanism accounted for in the simulation model?

Again, for the sake of simplicity, the locations of the actors’ power centers were set to opposite corners of a simulated, quadratic space. Moreover, the actors are assumed to have equal abilities to apply violence selectively at their power centers. The loss of accuracy is modeled as gradients that intersect at a fixed threshold of 0.5 which can be placed anywhere between the actors and is controlled by a single parameter: the territorial balance (TB). TB simply expresses the location of the 0.5 intersection of the accuracy values for the actors on a scale from 0 to 1. Values close to 0 put the simulated insurgents at an advantage: For most of the simulated space, they enjoy an advantage in accuracy. Values close to 1 express superior accuracy of the incumbent for most of the simulated space. Figure 4.1 illustrates the relationship between accuracy and territorial balance for different scenarios.

Mathematically, the relationship between territorial balance and accuracy is the following: TB must be set in advance by the user, such that $TB \in \{0, ..., 1\}$. The simulation runs on a two-dimensional grid and the distance of the diagonals is known and referred to as MaxDist in the following expressions. The Loss of Accuracy Gradient ($LAG$) for one side is then:

$$LAG = \frac{0.5}{TB \times MaxDist}$$

For each individual cell in the simulated space, the local accuracy for the
Figure 4.1: Illustration of the Loss of Accuracy Gradient in the ABM. Incumbent and insurgent have the ability to apply violence selectively in their power centers, but their ability to tell apart civilian bystanders and enemy combatants declines as the distance to their power centers increases. In the simulation, this decline is modeled as intersecting gradients. By definition, these gradients always intersect where both actors have a 0.5 accuracy, i.e. a 50% chance of hitting a civilian by accident.
cells at coordinates x and y can then be defined as the Euclidean distance to the actor’s power center at position (0, 0) multiplied by the LAG:

\[ Accuracy_{x,y} = 1 - (LAG \sqrt{x^2 + y^2}) \]

For larger distances from the power centers, the accuracy is truncated so that the minimal possible value is zero. Accuracy is then equivalent to the probability of hitting a civilian in an attack, since attacks are launched against one randomly chosen neighbor if the fraction of enemy combatants in the neighborhood exceed the accuracy value of the location of the agent. Clearly, this setup only calculates the local accuracy for one actor, but is easily repeated for the second one by swapping the power center positions and using 1-TB for the calculation of the LAG.

4.4.3 Reactive mobilization

Whenever civilian agents witness an attack on another civilian within their field of vision, they are alienated. The alienation factor also must be set in advance. Generally, the alienation factor is subtracted from the government loyalty of civilians whenever they witness an incumbent attack on civilians, meaning that the government is seen as less legitimate. It is added to their government loyalty whenever they witness an insurgent attack on civilians, which means that the government is seen as more legitimate in response to such attacks. For the sake of simplicity, if the government loyalty value falls below 0.2, the agents join the rebels. If it exceeds 0.8, they join the incumbent. Anything between 0.2 and 0.8 means that agents act as civilians with privately held loyalties that do not affect their behavior.

4.4.4 Running the simulation

With the local accuracy defined for the simulated space and the parameters for vision, government loyalty, and alienation defined, the simulation can be started. The main loop of the model executes a small number of steps defined
in pseudo-code in algorithm 4.1. For each agent, the following operations are repeated in each iteration of the main loop: Military actors can attack based on the fraction of enemy combatants within distance of their vision and the accuracy of their location. Civilians can witness attacks against other civilians within their vision. Attacked agents are killed and removed from the simulation and civilian witnesses of indiscriminate violence are alienated. The simulation stops as soon as no more military agents of one side are left.

\textbf{Algorithm 4.1} Pseudo-code for the main loop of the agent-based model.

\begin{verbatim}
for all military agents
   {if (accuracy of location < fraction of civilians in the neighborhood) \rightarrow attack
    if (enemy combatant hit) \rightarrow enemy combatant is removed
    if (civilian hit) \rightarrow civilian is removed and civilians in neighborhood are alienated}
for all civilian agents
   {if (government loyalty > incumbent threshold) \rightarrow civilian turns soldier
    if (government loyalty < insurgent threshold) \rightarrow civilian turns rebel}
for all agents
   {move one step in a random direction}
if (no military agents of one side left) \rightarrow stop
\end{verbatim}

While the logic of the simulation is rather straightforward, it is important to note that the exact scenario that results from the simulation is not deterministic. Instead, a wide variety of specific sequences of agent interactions can arise from different initial conditions and orders of execution. While following predefined distributions, the initial placement of the agents, the exact values of government loyalty, the order in which the actors execute attacks, as well as the order and direction of their moves are subject to random variability. Nevertheless, the introduced parameters and mechanisms generate strong probabilistic tendencies for the aggregate outcomes of the simulation. In the next section, a simplified example will illustrate how a probable sequence of agent interactions can arise from the defined parameters.
4.4.5 A simple example

In order to further familiarize the reader with the simulation, I will demonstrate the final steps of a simulated run manually. I am only going to focus on a very small section of the simulated space with a small number of agents to keep the effort tractable. This example, illustrated in figure 4.2, is intended to convey an intuition of how the model works to complement the algorithmic explanation above. The example represents a possible “end-game” situation: Close to the incumbent power center, a single insurgent, one incumbent, and three civilians interact in six steps. Three symbols are used to represent the political actors. The famous portrait of Ernesto Guevara represents insurgent forces, Hobbes’ Leviathan represents incumbent forces, and white flags represent civilians. Figure 4.2 illustrates a likely simulation scenario under superior incumbent accuracy.

For this manual exercise, I assume vision to be 1, meaning that agents can only see other agents in their immediate Moore neighborhood, i.e. the eight cells surrounding their own location. Moreover, I assume insurgent accuracy within this part of the simulated space to be 0, meaning that insurgents apply violence completely indiscriminately. Local incumbent accuracy is higher – say, 0.8. I also assume the civilian in the upper left corner of the space to be already siding with the incumbent, either due to initial loyalties or previous experience with indiscriminate insurgent violence.

In step a, insurgent and incumbent get a chance to attack another agent in their neighborhood. The incumbent refrains from attacking the civilian in its neighborhood because an attack would require a high fraction of its immediate neighbors to be insurgents. This fraction would have to be higher than the local accuracy, which I assume to be 0.8 in this case. The insurgent, however, attacks the only agent in its neighborhood, as it applies violence completely indiscriminately. As a result, the civilian in its neighborhood is removed from the simulation, as depicted in step b. This attack is witnessed by the civilian in the upper left corner of the simulated space. The civilian’s loyalty changes, meaning that its internal government loyalty is increased by one alienation.
This increase elevates the loyalty value above the defined cut-off for military mobilization. Therefore, the civilian turns into an incumbent agent in step c. All agents take one step in a random direction in step d. Conveniently for illustrative purposes, the incumbents close in on the remaining insurgent. In step e, the newly mobilized incumbent agent is allowed to attack first by chance. The fraction of insurgents in its neighborhood is 1, and therefore greater than the local accuracy, which leads the agent to attack. In step f, the last remaining insurgent is removed and the simulation run ends.

Clearly, many other scenarios could have developed from the initial conditions of the simulation. For example, the military actors could have attacked one another directly and any side would have succeeded with a 50% probability. Nevertheless, the discussed scenario was probable, given the advantage in accuracy for the incumbent. Exploring individual scenarios is an effective way of illustrating the functionality of the model, but it cannot communicate its substantive implications. In the next section, results from a large number of simulation runs are presented to show how the introduced micro-mechanisms of insurgency affect military outcomes and casualties on the macro-level.

### 4.5 Simulated macro-level predictions

As discussed above, in addition to conceptual clarity, simulation models also produce numerical predictions that can be compared to the empirical record. The purpose of such a comparison is to demonstrate generative sufficiency, i.e. the ability of the simulated micro-mechanisms to generate distinct macro-level outcomes or distributions of outcomes. To test for generative sufficiency, data from the running simulations and their final states is extracted for subsequent analysis. If the data is not found to correspond to basic attributes of the empirical record, the simulated theory must be rejected for its inability to account for the problem under investigation. If, however, the model does correspond to the empirical findings, the underlying theory must be considered a possible explanation.
Figure 4.2: An example of reactive mobilization in the agent-based model
To test for generative sufficiency, a Loss of Accuracy Gradient and reactive mobilization were built into the agent-based simulation. The substantive question that the simulation can help to solve is the following: If these mechanisms are really at work in insurgencies, how do they affect the military outcomes and approximate levels of casualties in insurgencies under varying geographic conditions? In order to answer this question, the territorial balance was systematically varied from 0 to 1 in 0.25 increments while parameters outside the theoretical focus of the study were set to reasonable values, meaning that initial loyalties were held neutral, most agents were initially civilians, and agents could see beyond their immediate neighborhood. For each parameter combination, 100 simulations were executed to generate a distribution of macro-level outcomes.

In figure 4.3, bold lines show the frequency of insurgent victories and the average fraction of casualties for each setting of territorial balance. The upper and lower bounds represent 1.96 standard deviation from the mean, which corresponds to a 95% confidence interval for normal distributions. The uncertainty in the simulated results arises from varying initial conditions, random execution orders, and the probabilistic nature of simulated combat.

As visible in figure 4.3, territorial balance strongly affects both military outcomes and casualties in civil wars. The model suggests that the ability of military actors to apply violence selectively strongly affects the aggregate outcomes in civil wars: Territorial balance is a crucial predictor of military outcomes in that it is negatively associated with insurgent success, as shown in figure 4.3. If both actors apply violence indiscriminately, repeated cycles of reactive mobilization on both sides lead to a more severe civil war. In this case, the effect of territorial balance on casualties is inverse U-shaped.

Of course, these predictions were generated under the assumption that violence against civilians actually leads to reactive mobilization, a claim generally not supported by deterrence-based accounts. Within the agent-based model,

---

2 Standard deviation for initial loyalties, alienation factor, and the standard deviation of initial loyalties were all set to 0.5. For systematic variations of these parameters outside the ranges discussed here, see section 9.1 on page 165.
alternative mechanisms are nevertheless easily simulated. Setting the alienation factor to 0 or even to negative values allows us to simulate civilian indifference or deterrence, i.e. support for the perpetrator in reaction to indiscriminate violence. Under such circumstances, the macro-level outcomes of insurgencies change. More precisely, the effect of territorial balance on military outcomes is reversed. For low TB settings, insurgent victory is extremely unlikely. Higher settings lead to higher probabilities of insurgent success. This reversal of the effect of territorial balance is easily explained: For low settings, the incumbent side will generally apply more violence, both selective and indiscriminate. This is because the steeper LAG for the incumbent makes the decision to attack in each round more likely. Regardless of whether civilians or enemy combatants are in range, incumbent forces will attack at a higher rate. If no reactive mobilization takes place for the insurgent side in response to indiscriminate incumbent violence, then higher levels of violence contribute to incumbent success. In this Hobbesian scenario, the quantity of violence makes all the difference while the quality of violence is more decisive under
Lockean assumptions. This insight is reflected in the numerical predictions of the agent-based simulation. The resulting effects are illustrated in figure 4.4.

Interestingly, changing the alienation factor does not affect the functional form of the relationship between territorial balance and casualties. For medium levels of territorial balance, the fraction of casualties is highest independent of civilian reactions to indiscriminate violence. It is important to mention, however, that the relative proportions of casualties differ: For positive alienation factors, the winning side usually suffers higher casualties, due to the fact that more civilians are mobilized for the winning side and the other side applies higher levels of violence. For negative or neutral alienation factors, the winning side suffers fewer casualties, as it is simply the more aggressive actor that wins the civil war. While this is a robust result of the simulation model, it cannot be tested across empirical cases of insurgency due to data limitations. Credible casualty counts are generally hard to come by, as I will discuss in chapter 7 and there is no data collection that disaggregates casualties by actor across cases of insurgency. Qualitative evidence suggests, however, that insurgencies are often successful even when confronted with heavy losses. Due to these data limitations, I decided not to systematically test for the opposite predictions of Hobbes and Locke with regard to the relative proportions of casualties in insurgencies. I will, however, perform empirical tests for the probabilities of outcomes and the overall levels of casualties, as well as direct tests of the assumed micro-mechanisms in later chapters.

How can these predictions be validated? Unfortunately, there is no general consensus in social science on how one should test agent-based models in empirical applications. Distributional properties of social systems have been replicated in simulation models (Cederman, 2010). Spatial and temporal predictions from simulation models have been compared directly to geo-referenced conflict events (Weidmann and Salehyan, 2012, Bhavnani et al., 2013). After reviewing these various options, I chose a rigorous yet easily communicable validation strategy consisting of a direct comparison of the average effect of the territorial balance in the simulation and the the average effect of an empirical
Figure 4.4: Simulated outcomes and casualties from 16,847 model runs. The bold lines show average results; the lower and upper bounds correspond to 1.96 standard deviation in the predictions. Note that the territorial balance was systematically varied from 0 to 1, while other variables were kept constant. In this case, a Hobbesian (deterrent) scenario with a negative alienation factor (-0.5) was simulated. The expected value of initial loyalties and the standard deviation of initial loyalties were again set to 0.5. Vision was set to 5 and the number of simulated agents to 600.

operationalization of the territorial balance that will be introduced in chapter 7. Naturally, such an average effect must be estimated by different technical means for the simulated and empirical insurgencies, but the underlying logic is the same in both cases. For estimating the simulated effects, it is sufficient to run a large number of simulations and record the overall level of casualties and military outcomes for each run. Keeping other parameters constant, the simulated territorial balance can then be varied systematically. Its average effect and confidence intervals can then be computed from the mean and standard deviation of the simulation runs for every setting of the territorial balance. Corresponding to the observed frequencies in the simulation runs, predicted probabilities from the econometric analysis can be calculated for various levels of territorial balance, while all other factors are held constant.
4.6 Discussion

In this chapter, I have introduced an agent-based model that communicates both the theorized micro-mechanisms as well as two observable implications on the macro-level regarding reactive mobilization in civil war insurgencies under both Lockean and Hobbesian settings. In the following two chapters, the empirical validity of the theorized micro-mechanisms – loss of accuracy and reactive mobilization – will be tested. After that, chapter 7 will compare the macro-level predictions of the simulation with results from a conflict-level analysis of 65 cases of insurgency.
5 The Loss of Accuracy Gradient

5.1 Introduction

In this chapter, I test the first empirical implication of the presented theory. The corresponding hypothesis (H1) states that the quality of violence in insurgencies declines as a function of distance from the attacker's power center. Empirically testing this and similar hypotheses has become easier in the past decade as the study of the micro-dynamics of violence in civil wars has gained strong momentum. Datasets encoding single conflict events, corresponding dates, and their exact geographic coordinates have been introduced (Raleigh and Hegre, 2005, Sundberg et al., 2011, Leetrau and Schrodt, 2013). Spatially referenced data that provide socio-economic context information to episodes of conflict are becoming increasingly available through public data platforms. This empirical development provides a strong basis for testing the presented expectations.

On the theoretical side, several explanations have been proposed for why military actors in civil wars apply violence selectively or indiscriminately. Kalyvas’ (2006) widely cited study brought wider scholarly attention to the topic, resulting in numerous publications (see for example Downes, 2007, Lyall, 2009, Lyall and Wilson, 2009, Condra and Shapiro, 2012).

Within this research program, two empirical questions have received increased attention: A series of studies seeking to explain under which con-

\[1\] See, for example http://growup.ethz.ch/ and http://data.geocomm.com/.
ditions civilians are intentionally harmed have focused on varying levels of military control (Kalyvias, 2006), initial motivations of the actors (Weinstein, 2007), and competition over local resources Metelits (2010). Other studies investigate the effects of indiscriminate violence on the conflict process and follow either hypotheses of alienation or deterrence resulting from the use of indiscriminate violence (Downes, 2007, Lyall, 2009, Lyall and Wilson, 2009, Linke et al., 2012).

Drawing on the simple distance-decay model proposed in section 3.3, this chapter shows that geographically operationalized power centers for both insurgent and incumbent are reliable spatial predictors of indiscriminate violence. Circumventing limitations in data availability, the analysis combines detailed insights into the war in Afghanistan with the wider coverage of a cross-conflict sample of violent events drawn from 10 additional cases of insurgency in African countries. The results clearly indicate that military actors apply more indiscriminate violence as the distance to their power center increases. The next section will review the existing literature in more detail. After that, theoretical expectations are tested in a large-N analysis.

5.2 Existing literature

The scholarly understanding of the driving forces behind indiscriminate violence has improved considerably in the last decade. Spearheading the recent turn toward the analysis of the micro-dynamics of civil wars, Kalyvias (2006, 69;149) convincingly argues that the scarcity of information and individual vulnerability in zones under enemy control leads to the application of indiscriminate violence. His theory of selective violence assumes a more complicated mechanism to be at work, however: Military actors use violence in zones of predominant but incomplete control to enforce collaboration and deter against defection. This leads to the empirical expectation that levels of selective violence should be highest in moderately contested zones of control, but not in areas of complete control or highest contestation. In an in-depth analysis of
the micro-dynamics of the Vietnam War, Kalyvas and Kocner (2009) show thatlocations of selective and indiscriminate violence tend to be separated spatially—a finding that lends support to their expectation that military control is a decisive factor. Focusing on the internal structure of military actors, Humphreys and Weinstein (2006) find that organizations that rely on material incentives to motivate their combatants and that lack the ability to punish indiscipline were more likely to apply indiscriminate violence. Along these lines, Weinstein (2007) shows that the behavior of military actors toward the civilian population is affected by their initial motivations: Ideologically motivated rebels are less likely to engage in acts of indiscriminate violence than materially motivated ones. Wood (2010) points to the discriminatory capacity of rebel organizations, arguing that weaker rebel organizations might lack the capacity to discriminate between combatants and civilians. Goodwin (2006) argues that past interactions between rebels and civilians affect future episodes of violence. Offering yet another perspective, Metelits (2010) analyzes how violence becomes more indiscriminate when military actors compete over control and resources within the same territory.

A noteworthy aspect of these theories is their focus on factors endogenous to the changing levels of military capabilities, such as territorial control, effective policing, and the availability of information about civilian loyalties.

While this perspective has generated many important insights, fully endogenous explanations provide a bleak outlook for informing humanitarian relief operations or political decision makers: Instead of being able to identify regions that face a higher risk of being affected by indiscriminate violence ahead of time, such explanations suggest that closely following the conflict process is the only reliable way to predict where indiscriminate violence is likely to occur next. As discussed in section 3.3, I provide an exogenous explanation for why actors are more likely to use indiscriminate violence far away from their power centers. While not disputing the role of military control, initial motivations, civilian-military relationships, and conflict intensity, I expect the fundamental inability of military actors to apply violence selectively outside
their power centers to strongly affect patterns of violence in the statistical aggregate. The next section will articulate specific empirical expectations based on the presented theory.

5.3 A distance-decay model

To quickly recapitulate the theoretical discussion, I assume that violence declines in selectiveness as a function of distance from the actors’ power centers. Modifying Boukling’s (1962) “Loss of Strength Gradient” to a “Loss of Accuracy Gradient” (LAG), I therefore assume the quality of violence deteriorates as a function of distance and not primarily its quantity. Within this model, the initial levels of accuracy are not necessarily the same for all actors and the gradient in the declining quality might also vary. For example, modern means of surveillance or a better intelligence apparatus might very well lead to an increased ability of the actors to identify enemy combatants and collaborators, but this ability to apply violence accurately is expected to decline with growing distance to their power centers.

Operationalizing power centers geographically is necessary to test this hypothesis. For the state, the capital city naturally qualifies as its center, as it is the physical location of leading decision makers and the highest echelons of the military. This presumes, of course, that the peripheral insurgency and the central government represent the central conflict dyad. Regional actors that represent the government in the periphery might have additional strategic locations to operate from. In most insurgencies, however, capturing the capital city is also the final objective of the uprising and extending central rule into the periphery is the final objective of the state. For the rebels, the situation is less obvious. McColl (1969) describes several determinants for the emergence of rebel bases in irregular conflict, such as distance to international borders and terrain accessibility. While controlling for these factors, I assume the rebel power center to generally emerge in areas most remote from the capital city.

Mao’s “Protracted War” model explicitly assumes that fighting in the initial
stages of an insurgency takes place in the periphery. Once the more remote
gerions are secured, rebel forces advance toward the capital city in multiple
stages of their campaign. The communist insurgencies in China, Cuba, and
Vietnam, as well as the anti-Soviet insurgency in Afghanistan all reflect this
pattern. This allows for the simplifying assumption that the center of state
power can be associated with the capital city and that the rebels’ realm is the
periphery. This assumption is in line with the communist literature (Guevara,
1961, 10), counterinsurgency studies (Galula, 1964, 23-24), recent conflict re-
search (Kalyvas and Balcells, 2010, 415), and agent-based simulation studies
(Cederman, 2008). Empirical evidence also indicates that rebels indeed seek
out the most remote regions to start insurgencies. Macaulay (1978, 288) rep-
ports that the Cuban 26th of July Movement operated from the easternmost
province of the island – the Sierra Maestra mountains – and then gradually
moved toward Havana. Nolan (1958, 71) observes that “in the twentieth cen-
tury, those seeking power for the purpose of radically transforming society have
generally turned to rural-based guerrilla warfare as a means of overthrowing
the existing order”. Chalk (2008, 5) summarizes the initial phases of the in-
surgency in southern Thailand with an emphasis on its geographic scope: “the
main aims were to present the southern provinces as an area that remained
beyond the sovereign control of Bangkok”. Using international borders for re-
treat and supply, insurgent movements tend to use remote areas to build up
their bases (Salehyan, 2009, Hironaka, 2005, 76).

For these reasons, I operationalize the insurgent power center as the area of
a war-torn country that is most remote from the capital city while still being
affected by conflict. Clearly, such assumptions about the locations of rebel
strongholds cannot be made for all civil wars. For example, the Yugoslav civil
war split the country mainly along ethnic lines and not in terms of a territorial
divide that arose from a popular rebel movement challenging the state. In
such cases, the pre-war settlement locations are a much better predictor of
violence than distance to the former capital city (see Weidmann, 2011). But
the theoretical scope of this study are insurgencies, i.e. civil wars in which
irregular insurgents draw on guerrilla tactics to weaken the regular army of the incumbent and generate civilian support. Such conflicts are waged from the periphery in the majority of cases. Moreover, insurgents utilize difficult terrain in surprise attacks to compensate for the military superiority of the state. Based in these considerations, the empirical tests of hypothesis 1 draw on the distance to the capital city as the main independent variable. More precisely, the following sub-hypotheses will be tested:

**H1.1:** Distance to the capital city has a positive effect on incumbent indiscriminate violence.

**H1.2:** Distance to the capital city has a negative effect on insurgent indiscriminate violence.

**H1.3:** Areas closest to and farthest away from the capital city see the highest levels of indiscriminate violence.

Moreover, several control variables are included in the empirical analysis that express terrain accessibility in terms of natural landcover, elevation and distance to the next major city. By focusing on exogenous conditions that affect the spatial distribution of indiscriminate violence in civil wars, the approach does not fundamentally challenge the insights generated by others, but it allows for central components of their theories, such as Kalyvas' (2006) “identification problem”, to be associated with geographic conditions. Moreover, instead of having to systematically analyze patterns of control or rebel-civilian relations in wartime across conflict to acquire generalizable insights, this approach lends itself to a comparatively robust and transparent empirical design. Based on modest assumptions, the predicted association between the geographic configuration of a theater of war and the predominant types of violence that occur can be tested. The next section discusses the testing procedure.
5.4 Data and case selection

The increasing availability of conflict events data has generated unprecedented possibilities for analyzing the subnational characteristics of civil wars. Such data have been used in a large and rapidly growing number of publications on the micro-dynamics of civil war (Hegre et al., 2009, Schutte and Weidmann, 2011, Zammit-Mangiona et al., 2012, Weidmann and Salehyan, 2012, Braithwaite and Johnson, 2012). While most of these studies focus on single conflicts, a series of studies have also attempted to find generalizable patterns in events data across conflicts, such as the local determinants of conflict intensity and the relative locations of primary conflict zones (Buhaug and Rød, 2006, Buhaug, 2010). However, a decisive factor sets these studies apart from the empirical analysis of the introduced Loss of Accuracy Gradient: Instead of modeling the presence or absence of violence, this theory focuses on the type of applied violence. This distinction calls for an empirical record that reveals the quality of violence as well as its location. Moreover, acts of violence must be attributable to one of the military actors in order to test the proposed distance-decay mechanism. Unfortunately, such data is hard to come by.2

A conflict events dataset covering African civil wars between 1990 and 2010 has been released that provides information on both civilian and military casualties (Sundberg et al., 2011). The “Georeferenced Event Dataset” (GED) focuses exclusively on lethal encounters in civil wars. Counts of civilian and military fatalities are given for each conflict event. A caveat for the problem at hand is that conflict events cannot be attributed to one of the military actors. This is understandable from a data collection point of view as the coding relies on media sources. For journalists, competing claims about the initiation of violence are often hard to verify and the affiliation of the perpetrators is usually the most contested part of any story emerging from the turmoil.

---
2The Armed Conflict Location and Event Dataset (ACLED) (Raleigh and Hegre, 2005) provides geo-coded information on acts of violence, but it does not differentiate between civilian and military casualties. While certainly useful in other cases, these data do not lend themselves to the analysis of variations in the quality of violence as a function of location.
of civil war. Consequently, the GED does not code the initiator of violence. Beyond the realm of basic research, data on conflict events have also been collected by armed forces engaged in counterinsurgencies. The most recent data collection system for the micro-dynamics of ongoing conflicts within the US military is called SIGACT, an abbreviation for “significant activity”. SIGACT files are referenced in time and space, indicate the specific type of incident, and record casualties and the initiator of the event. As such, SIGACT data is not intended for release to the general public, but the data collections for Afghanistan were made available via wikileaks.org on July 27, 2010.  

The SIGACT data provides all necessary information for testing the proposed hypothesis. SIGACT is nevertheless restricted to two cases – Afghanistan and Iraq – and only Afghanistan clearly qualifies as an insurgency, while the war in Iraq blends elements of an insurgency with ethnic and communal violence (see Weidmann and Salehyan, 2012).

Given these restrictions in data availability, I decided to combine the detailed insights that can be gained from SIGACT with the larger coverage of the GED in two separate empirical analyses. In a first step, I conduct an analysis of the SIGACT data from Afghanistan. In this case, the initiator of violence is clearly coded. However, the casualty figures from SIGACT are not trustworthy, as soldiers in the field have strong incentives to downplay the levels of civilian casualties resulting from their actions. Moreover, confirming casualties is easier in certain scenarios than in others. For example, calling in an airstrike against a suspected enemy position might very well result in civilian casualties that go unnoticed. As a result, a related study on civilian casualties in Iraq relied on external casualty counts (Condra and Shapiro, 2012). To circumvent this problem, I operationalize selective and indiscriminate violence within the SIGACT study not in terms of the alleged casualties,

---

3Using this data in the subsequent empirical analysis was not an easy decision. I am using the data under the premise of not revealing the activities or identities of individuals involved in the conflict in Afghanistan. Moreover, I do not analyze which side is responsible for higher levels of civilian casualties. Instead, I am using the leaked data to generate insights for basic research by aggregating available information, not exposing details. I therefore decided that the benefits for the scholarly understanding of conflict would outweigh the risks associated with analyzing this unique dataset.
but in terms of the type of violence employed by the actors, very much along
the lines of the theoretical discussion on the technological constraints of selec-
tive violence. After that, I test for the validity of the theory across conflicts in
an analysis of the spatial distributions of civilian and military casualties based
on the GED. In summary, the empirical strategy seeks to combine the detailed
insights of SIGACT data with the wide coverage of the GED.

5.4.1 Explanatory variables

For the samples of observations from SIGACT and GED, I coded explanatory
variables through a GIS procedure. Distances to the power centers were estab-
lished based on the CShapes dataset (Weidmann et al., 2010). For Afghanistan,
I included a distance measure to the Pakistani border, as transborder insurgent
activity from the Pakistani tribal areas accounts for a substantial amount
of SIGACT recorded events.

To account for other factors beyond these main independent variables, I
coded the approximate terrain elevation from a global elevation model (Gesch
et al., 1999). Higher elevation is associated with rough terrain that is less
accessible to the state (see Fearon and Laitin, 2003), which in turn gener-
ates incentives to apply violence more indiscriminately in such regions. Approx-
imate population counts from CIESIN (2005) were also coded for each
conflict event. The effect of population concentrations on the spatial distribu-
tion of conflict events within countries is well researched (see Raleigh and
Hegre, 2009). Arguably, population concentrations also affect the quality of
violence: In less populated areas, military actors single out enemy combat-
ants more reliably than in densely populated cities. For the SIGACT study,
the number of previous conflict events within a 50-kilometer radius less than
one month prior to the incident was also coded. This variable was included
to account for the effects of conflict intensity on the quality of violence. If
actors are engaged in prolonged and intense stand-offs, they might be more
inclined to resort to indiscriminate violence. Furthermore, measures of wealth
from a spatially referenced dataset on the subnational distribution of goods
and services was included (GECON) (Nordhaus et al., 2006). As with population concentrations, correlations between spatially disaggregated wealth and the geographic distribution of conflict events (Hegre et al., 2009), as well as poverty and initial fighting in civil wars (Buhaug et al., 2011) have been found.

With regard to the types of violence used by the actors, the economic value of the conflict side should matter as well: Incumbent forces could resort to more indiscriminate violence to defend such strategically important locations at all costs. Insurgents could resort to indiscriminate violence to disrupt operations at such locations. As an additional control, I added an estimate for the natural landcover (Hansen et al., 2000) at the site of the conflict event. The rationale behind adding this variable is that densely forested areas are also less accessible to the state. As a result, such areas could be seen as rebel strongholds and attacked based on indiscriminate tactics. A measurement for the effective traveling distance to the nearest city that had at least 50,000 inhabitants in the year 2000 (Nelson, 2008), hereafter referred to as urban distance, was also taken into account. Major cities tend to be under incumbent control at least in the early stages of irregular insurgencies. Insurgents therefore lack the means to apply violence selectively in major cities.

Finally, I constructed a line-of-sight measurement to account for a tactical particularity of the conflict touched upon above: In areas with limited lines-of-sight due to high densities of natural obstacles, actors might rely on indirect fire instead of direct fire. Since I coded the first type of attack as selective and the second type as indiscriminate, I needed to control for this factor. The line-of-sight measurement was constructed in several steps: Using the digital elevation model by Gesch et al. (1999), I calculated the number of surrounding cells that are visible from any location. This calculation involved the specification of “horizon” in terms of a maximal distance from the cell under investigation (30 kilometers in this case). For all cells within that horizon, Bresenham’s (1965) algorithm was used to calculate all cells along a straight line connecting the origin and the target cell. In a second step, elevation levels along this line were used to calculate angles between the cell under investigation and
the cells along the line. The number of visible cells was then established by counting the number of cells along this line for which no steeper angle had been calculated for any preceding cell. In this way, cells along this line with no obstructing cell in front of them were established. Since this procedure was repeated for all cells within the horizon, a count of all visible cells was established for each cell in the elevation dataset. The explanatory variables were available as geo-referenced data with varying resolutions. In order to associate the explanatory variables with the conflict observations, I mapped the conflict events to their nearest geographic neighbors in the explanatory datasets. As a result, samples from GED and SIGACT could be analyzed in a multivariate regression analysis.

5.5 The war in Afghanistan

The ongoing war in Afghanistan was identified as a typical case of insurgency which lends itself to testing the more general hypotheses discussed above. Afghanistan is a typical case of insurgency due to its general socio-economic conditions, the composition of the uprising, the international context, and the sequence of events at the macro-level, which are all typical of this type of conflict. This makes Afghanistan a suitable case for testing the more general theory presented in chapter 3 (on this type of case selection see Seawright and Gerring, 2008, 297).

From a socio-economic point of view, the country is a risk candidate for civil war. Widespread poverty, a weak central government, a recent history of intense political violence, forbidding mountainous terrain, and a patchwork of intermingled ethnic groups with varying access to political power and wealth make Afghanistan a prime candidate for civil war with regard to the central variables associated with war onset (Collier, 2000, Buhag and Gates, 2002, Fearon and Laitin, 2003, Cederman et al., 2010, 2011b).

The composition of uprising is almost prototypical for conflicts of this kind, combining an irregular local insurgency with the build-up of a shadow ad-
ministration. Moreover, clandestine international support for the insurgency is assumed to take place and can be found across a variety of cases, such as Chechnya and Ingushetia (Moore and Tumulty, 2008), Vietnam (Sheehan, 1988, 650), and, of course, the anti-Soviet insurgency in Afghanistan (Wright, 2007, 120). Apart from that, the sequence of macro-events that led to a large-scale insurgency is also typical of a wider class of cases: A government is replaced through outside intervention and subsequent occupation of the country. The new government faces a problem of legitimacy and is heavily reliant on outside support. Elements loyal to the former administration start a protracted campaign to topple the new incumbent.

5.5.1 SIGACT data

The version of SIGACT used here covers the time period from 2004 to 2010 and amounts to 76,247 records. All records are time- and geo-referenced, coded as insurgent or incumbent activity, and distinguish between 154 types of events. SIGACT reports are passed up the chain of command, sometimes from the platoon level, allowing for an extremely detailed view of the conflict from the U.S. perspective. From a quantitative standpoint, the SIGACT data provide the most complete view on the Afghan war. As mentioned before, one shortcoming of this data is that it was coded by soldiers in the field that perceive the conflict subjectively and report violence events with strong incentives to attribute civilian casualties to the enemy, or to not report them at all. Therefore, I decided to operationalize indiscriminate violence as the use of heavy arms, as described in detail below.

A cross-sectional view of the data reveals the existence of two major conflict zones as illustrated in figure 5.1. One zone centers around the city of Kandahar in the Helmand province while the other lies along the mountainous tribal areas bordering Pakistan. The northern and eastern parts of the country seem less affected by civil war, although some instances of violence are recorded in SIGACT.
Figure 5.1: The figure above shows a (Gaussian) density estimate for the locations of conflict events in Afghanistan between 2004 and 2010. It gives an impression of the relative intensity of conflict in different parts of the country. Violence was most intense in Helmand province and in the tribal areas bordering Pakistan. The spike in the northeastern part of the country is the result of intense fighting in the Korangal valley.

5.5.2 Operationalizing indiscriminate violence in SIGACT

The classification of violent events in this section follows the premise of section 3.1: Based on the discussion of the technological determinants of the quality of violence, I assume certain classes of incidents in SIGACT to be less selective than others. I focus on conflict events that are especially frequent in the empirical record to allow for the largest possible external validity of the analysis and to prevent the insights from being driven by marginal conflict episodes. Moreover, these event types can be easily identified as being more or less selective based on the discussion of the technological prerequisites of selective violence discussed in section 3.1.1. The following overview specifies the corresponding event categories and justifies the coding choices. Moreover, the selection of relevant events focused on types of events that are comparable for incumbent and insurgent. Table 5.1 gives an overview of the event categories associated with incumbent and insurgent violence. In the following subsections, I will discuss the coding choices in more detail.
Table 5.1: Event categories for selective and indiscriminate violence in the SIGACT data

Indiscriminate insurgent violence

Mine strikes were counted as acts of indiscriminate violence. In order for an explosive device to classify as a landmine it must be victim-activated. It is this technical particularity that renders landmines extremely inaccurate, since no identification of the target is possible by the attacker. Similarly, indirect fire allows an attacker to hit targets beyond his line of sight. Moreover, it allows the attacker to deliver heavy and explosive munitions over greater distances. These tactical characteristics come at a decisive cost in population-centric warfare: limited accuracy and high lethality. Effective indirect fire usually relies on an artillery spotter having line-of-sight contact with the target to report back to the shooter. It usually requires several iterations of shooting and re-aiming to hit a target. More importantly, explosive munitions destroy the homes and property of innocent bystanders, even if they do not physically harm civilians. These characteristics make indirect fire less discriminate than direct small arms fire and these events were therefore counted as indiscriminate.

Selective insurgent violence

Direct fire by insurgents has certainly claimed the lives of civilians, but it still provides a more selective way of targeting collaborators and incumbent forces than indirect fire. More importantly, due to line-of-sight contact being a precondition of the use of direct fire, insurgents at least know what or whom they are shooting at in combat situations.
**Indiscriminate violence by US Forces**

The rules of engagement of US Forces put restrictions on using lethal force. Although ISAF can probably use *indirect fire* more professionally than insurgent forces, the problem of high lethality combined with low accuracy remains. Even if measures are taken to spare the lives of bystanders, large-scale material destruction is still a natural byproduct of explosive munitions. Therefore, instances of *indirect fire* were counted as indiscriminate violence for the incumbent side. Generally, SIGACT does not contain information on Air Force activity in Afghanistan. This is due to the fact that soldiers on the ground file the reports and only sometimes include references to air strikes. If air strikes were carried out as part of other fighting activities, the incident might simply not be labeled as such. Nevertheless, *close air support* was counted as an instance of indiscriminate violence since it applies violence more destructively and less selectively than direct small arms fire.

**Selective violence by US Forces**

*Direct fire* was also counted as an instance of selective violence. Again, these events are frequent enough to allow for generalizable insights and arguments for counting direct fire of the insurgent side as selective; this also applies to the incumbent side.

**5.5.3 SIGACT results**

To establish the effects of the distance to the power centers on the type of violence applied, I estimated logistic regression models for both insurgent and incumbent violence. Corresponding results can be found in table 5.2 on page 94. The unit of analysis in these models is the conflict event. I reduced the sample of all SIGACT events to those that relate to violent incidents coded either as selective or indiscriminate according to table 5.1 on the facing page. Models 1, 2, and 3 predict indiscriminate violence for the insurgent side while models 4, 5, and 6 predict incumbent violence. Models 2 and 5 contain all explana-
tory variables, whereas 3 and 6 consist of subsets of the explanatory variables chosen to improve the goodness-of-fit of the models as expressed in relatively lower AIC values. AIC is a suitable statistic for expressing the benefits of adding explanatory variables to statistical models against the associated costs: More explanatory variables usually increase the likelihood of the model, but they can also lead to overfitting. By penalizing for the number of explanatory variables and rewarding likelihood, AIC provides a simple relative metric for comparing statistical models (Akaike, 1974). Models 1 and 4 do not contain the main independent variables (the distances to Kabul and the Pakistani border) and serve as a baseline for the AIC statistic. Comparing models 1 and 2 with regard to their AIC scores therefore allows us to see that the distance variables substantially improve the model fit instead of merely leading to overfitting. Similarly, a comparison of models 4 and 5 shows that the inclusion of the distance variables also lowers the AIC for the incumbent side, which further underlines the relevance of the variables.

In summary, the lower AIC values indicate that the distance variables substantially improve the model. They do not, however, show how the explanatory variables are associated with types of violence used by the actors. I will therefore discuss the results from a substantive point of view below. As first glance, table 5.2 shows a number of variables significantly associated with the quality of violence used by the actors. These very strong results are certainly at least partially driven by the large number of observations – 22,512 events for the insurgent side and 1,288 events for the incumbent. But more importantly, the fact that these geographic variables are significant predictors of the type of violence indicates that locations of selective and indiscriminate violence tend to be separated geographically, which is in line with the theoretical expectations (see also Kocher et al., 2011).

*Distance to Kabul*, the main independent variable, is negatively associated with insurgent indiscriminate violence, but positively associated with incumbent indiscriminate violence. This confirms hypotheses 1.1 and 1.2 introduced above: Close to the capital city, insurgents lack the ability to discriminate
between combatants and bystanders and are confronted with tactical and cognitive incentives to apply violence indiscriminately. For the incumbent, this situation is reversed: Selective violence is more likely to be applied close to the capital and deteriorates in quality as distance increases. The magnitude of these effects can be seen in figure 5.2 which shows the predicted probabilities for the different actors and the different types of violence.

The results for the elevation variable is negative in models 2, 3, and 4, but positive in model 1 and not significant in model 5. It is therefore probably not advisable to read too much into these estimates. A possible cause for the negative estimate for both sides is that the use of heavy arms is more difficult in the mountains, but as stated above, this result is not conclusive and should be taken with a grain of salt. Population has a positive effect on insurgent indiscriminate violence, but a negative, non-significant one for the incumbent. This result likely reflects the predominantly incumbent control of cities and larger hamlets. Such areas are densely populated and controlling them is a strategic objective of the state. Policing such relatively confined areas is also usually possible for the incumbent. Such a pattern of control has been described as typical in the literature (Sheehan, 1988,50; Kalyvas, 2006, 134). The line-of-sight variable is only significant in model 1, and is therefore difficult to interpret. A strong predictor of the type of violence applied, however, is GECON – the spatially disaggregated measure of wealth. Interestingly, the variable is negatively associated with insurgent indiscriminate violence, but positively associated with incumbent indiscriminate violence. As argued above, a possible explanation for this effect might be that incumbent forces use more heavy weapons in order to secure and defend valuable areas and infrastructure to prevent them from falling into the hands of insurgents. An alternative explanation is that much of Afghanistan’s wealth is located in the border region to Pakistan, in terms of natural resources (Peters et al., 2007). Following hypotheses 1.1 and 1.2, this region should see higher levels of incumbent indiscriminate violence and lower levels of insurgent indiscriminate violence. Further analysis would be required to fully disentangle these two
possible explanations.

_Urban distance_ has a positive effect on insurgent indiscriminate violence and no significant effect for the incumbent side. At first glance, this result is contrary to the expectation that insurgents use indiscriminate violence in the cities. However, the specific coding of the variables associates indiscriminate violence with the use of heavy weapons, such as mortars. Therefore, the result is most likely driven by the fact that insurgents often refrain from using mortars and mines in the cities, restricting their use to more rural settings which offer better opportunities for retreat after an attack. This tactical particularity is less likely to drive the estimates on the incumbent side, as incumbent forces can rely on long-range artillery and air support and do not need to retreat after every engagement.

The positive effect of natural landcover on insurgent indiscriminate violence can be interpreted along these lines as well: Insurgents lack the means to transport heavy weapons in the open. Given the efficiency of modern air surveillance, it might be necessary for insurgents to utilize landcover for transporting these weapons. Again, due to the clustering of insurgent activity along the Pakistani border, local conditions can drive the statistical inference. With many parts of the country sparsely forested, the Nangarhar province in eastern Afghanistan is an exception. With a comparatively mild climate and a continuous fresh water supply from the Hindu Kush, the region is comparatively forested and engaged in timber production. However, it has also seen increased insurgent activity due to its geographic proximity to the border. Again, additional testing would be required to fully rule out this alternative explanation for the effect.

The _previous violence_ variable expresses the number of conflict events that took place up to one month prior to the event under investigation and within a 50-km radius. Interestingly, the estimate for the incumbent side is negative, while it is positive for the insurgent side. Asymmetry in the applied tactics is a possible explanation for this effect: Insurgents tend to commit “hit and runs”, i.e. leave the area after initiating an attack, while incumbent forces hold
on to territory. As the battle progresses, the incumbent side increasingly relies on air and artillery strikes to defend its positions, while the insurgent side relies more on mobile attacks. In summary, the variables *distance to Kabul*, *GECON*, *landcover*, and *previous violence* have opposite effects on incumbent and insurgent indiscriminate violence. The results for the main independent variable is in line with the theoretical expectations. While these results give a first indication of the explanatory power of the introduced theory, they only reflect the conditions within one single conflict. In order to test the theoretical expectations more generally, I also conducted an analysis of irregular civil wars in Africa which will be discussed in the next section.

5.6 Civilian casualties in African insurgencies

To broaden the analysis beyond Afghanistan, I also analyzed a subset of the conflict events from the Georeferenced Event Dataset (GED). An overview of these cases is given in table 5.3. The GED analysis was restricted to cases of irregular war and identified through a dataset of insurgencies by Lyall and Wilson (2009). As mentioned previously, the wider coverage of the GED data comes at a price: Initiators of violence are not coded, and the type of violence for the conflict events is not disaggregated into tactical categories. As a result, the dependent variable of this analysis is not the type of battle event, but the number of civilian casualties that resulted from the application of violence.

Bearing in mind the theoretical discussion, the empirical expectations of this section are straightforward: The two actors employ violence indiscriminately close to the other actor’s power center (H1.3). In terms of casualties, this would imply that civilian casualties are most likely to arise in two locations: close to the capital city where insurgent forces are more likely to engage in indiscriminate actions, and in the most remote regions where incumbent forces employ violence indiscriminately. This theoretical expectation translates into a U-shaped effect for the number of civilian casualties as a function of distance between the capital city and the nearest international border.
Figure 5.2: Plots showing the effect of the main independent variable on the probability of indiscriminate violence for both incumbent and insurgent. To generate these figures, *distance to Kabul* was systematically varied while all other variables were held constant at their means.
5.6.1 GED data

The GED is a collection of detailed data coded and maintained by Uppsala University. Covering lethal events from both civil wars and communal unrest, the GED provides 20,396 observations of violence that took place between 1990 and 2008 in Africa. The GED data is based on an elaborate coding procedure that ensures reliability by cross-validating records with multiple coders (Sundberg et al., 2011). Most importantly for this study, the GED provides geographic coordinates along with all observations, allowing the hypotheses derived above to be tested empirically.

Since the theoretical scope of the argument is restricted to insurgencies where rebels usually operate from the periphery, a subset of conflicts in the GED had to be identified that meet this criterion. I therefore used a dataset by Lyall and Wilson (2009) that is restricted to clear-cut insurgencies. Unfortunately, this dataset and the GED only overlap partially, since the GED codes conflict events on the African continent that took place between 1990 and 2008, while the Lyall and Wilson (2009) dataset codes military outcomes of insurgencies on a global level from 1800 to 2010. As a result of this partial overlap, conflicts from only 10 countries appear in both data collections. The resulting sample consists of 3,744 observations from the GED. This sample was generated by preserving only those GED observations that took place during and within the same country as the insurgencies coded by Lyall and Wilson (2009). Moreover, instances of communal violence and one-sided violence were excluded from the sample since their occurrence is beyond the scope of the theory. Distances to the capital city were normalized for each country, using the most remote conflict event as the maximal possible distance. All observations were pooled and analyzed in the aggregate.

4In a robustness check, I also ran the analysis on the unrestricted sample including all GED events and found that the substantive results also hold for the entire sample (see section 5.6.3). This test was necessary to show that I am not selecting cases that match the theory while omitting others.
Figure 5.3: The figure above shows (Gaussian) density estimates for the locations of conflict events in the ten GED cases under investigation. Conflict events tend to cluster close to the capital cities and in the most remote regions.
5.6.2 Operationalizing indiscriminate violence in GED

With regard to the GED, the quality of violence is more easily measured than in SIGACT: Since every observation provides counts for military and civilian casualties, the numbers of civilian casualties directly reflect higher levels of indiscriminate violence. Therefore, the dependent variable in the GED study is the number of civilian casualties in violent events, while the dependent variable in the SIGACT study is the type of applied violence according to table 5.1.

5.6.3 GED results

In a second empirical analysis, conflict events from 10 cases in the GED were analyzed (Sundberg et al., 2011). Corresponding results can be found in table 5.4 on page 96. Although they do not reveal the initiator of violence, the GED observations code civilian and military casualties. Predicting the number of civilian casualties as a function of location allows us to test the proposed theory from another angle and with a considerably wider scope.

Model 7 presents only the main independent variable, and model 8 only the additional controls. Model 9 contains all explanatory variables and model 10 only those that contribute to a lower overall AIC score. Model 11 was run with the full sample of all civil war-related GED events, excluding only interstate war and one-sided violence. This model was included as a robustness check to rule out the possibility that the central insights were driven by the case selection. Model 12 contains a smooth term for the main independent variable. A first indication of the relevance of the normalized distance variable is the fact that the full model 9 features a lower AIC score than model 8, which contains all but the main independent variable. Since indiscriminate violence cannot be attributed to one specific actor in the GED, only the expectation that the capital city and the periphery should see the highest levels of indiscriminate violence can be tested. Hypothesis 3 expresses the corresponding empirical expectation of a quadratic (U-shaped) effect for the normalized distance from
the capital city. As visible in table 5.4 on page 96, the quadratic specification of the normalized distance to the capital is positively and significantly associated with the number of civilian casualties in models 7, 9, 10, and 11. Correspondingly, the smooth estimate in model 12 for the distance to capital variable roughly corresponds to this U-shaped form (see figure 5.6.3). Again, the strength of the effect is most easily shown in predictions that systematically vary the distance variable while keeping all other variables constant at their means (figure 5.6.3).

The control variables for the GED were the same as in the SIGACT analysis for Afghanistan, except for previous violence, which was not coded due to the relative scarcity of reported events. Moreover, the number of military deaths is coded for each GED event and was included as a control variable. What effects can be expected for the additional control variables? Given the very large share of insurgent events in the SIGACT sample (94.2% of all observations) and the overwhelming qualitative emphasis on the fact that rebels tend to initiate more attacks in insurgencies, one can assume that most of the recorded events in the GED were triggered by insurgents. Therefore, the sample of GED events should mainly consist of insurgent attacks and the directions of the effects should correspond to those of the SIGACT analysis of insurgent indiscriminate violence. As in the previous analyses, the logged population count has a positive estimate, but no significant effect in the full model. A strongly significant and consistently negative effect can be found for the urban distance variable which had a positive effect on insurgent indiscriminate violence in the SIGACT analysis. The fact that GED is based on media accounts could drive this result at least partially. Arguably, events in and around large cities are more frequently reported in international media, as the journalistic coverage of rural areas is logistically more difficult. Natural landcover is only significant in the baseline model without the main independent variable, but the corresponding estimate is positive, and the results seem too weak to be safely interpretable. The GECON variable is again negatively associated with higher levels of indiscriminate violence, which corresponds to the SIGACT
models for insurgent activity in Afghanistan. As argued above, military actors tend to fight over economically valuable locations. Multiple violent encounters at those locations are likely to lead to military casualties, but not necessarily to civilian casualties, as civilians can avoid these locations.

Contrary to the SIGACT analysis, line-of-sight is positive and significant for all estimated models. This implies that lowlands are generally more likely to see high levels of civilian casualties than mountainous regions. The high lethality of heavy arms, such as tanks, that cannot be used in forbidding terrain can cause this effect. The military deaths variable is also positive and significant for all models estimated for the 10 cases of insurgency, but negative for the unrestricted sample, which is surprising. The positive estimate corresponds to the intuition that higher levels of military casualties also lead to higher levels of civilian victims.

5.7 Discussion

The empirical analysis has confirmed the theoretical expectation that locations of violent encounters strongly predict civilian casualties as well as the means of violence used by the actors. Focusing centrally on the distance to the incumbent power center, the SIGACT analysis has shown that indiscriminate violence as operationalized through the use of heavy arms is most likely to be applied close to the capital city by insurgents, but far from the capital by incumbent forces. A corresponding analysis with a wider empirical scope has shown that both remote and central regions see the highest levels of civil casualties while controlling for military casualties. This result holds for the restricted sample of clearly identified insurgencies, but also for the full GED sample of civil war violence in African conflicts between 1990 and 2010.

Most importantly, these results suggest that patterns of indiscriminate violence in civil wars can be predicted ahead of time. Instead of being fully endogenous to levels of military control or contestation, initial motivations, or competition over resources, some variation in the quality of violence can be
explained exogenously. Empirical support for this claim comes from a cross-sectional analysis of a larger number of cases; subsequent research might be able to model areas that are prone to indiscriminate violence more accurately. With regard to the wider scope of this study, the empirical reality of a “Loss of Accuracy” as a function of distance has been confirmed. The next chapter will discuss methodological challenges and a possible solution for testing the second empirical implication of the presented theory: reactive mobilization in response to indiscriminate violence.
Figure 5.5: Smooth estimate for the effect of distance to capital in the GED analysis. Note the U-shaped appearance with a predominantly negative effect in the middle and a positive estimate for very large and very small distances to the capital. The positive spike at around 0.6 is significant, but weaker than the effects at minimal and maximal distances, as well as the negative estimate at 0.4.
Table 5.2: Regression results from the SIGACT analysis. The estimated models predict indiscriminate violence as a function of

<table>
<thead>
<tr>
<th>Distance to Rel. (km)</th>
<th>Distance to K memb (km)</th>
<th>Landcover</th>
<th>Urban distance</th>
<th>Population</th>
<th>Previous violence</th>
<th>Previous violence</th>
<th>Previous violence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: ∗∗∗ represents p < 0.001, ∗∗ represents p < 0.01, ∗ represents p < 0.05.
<table>
<thead>
<tr>
<th>No.</th>
<th>Country</th>
<th>Scenario description</th>
<th>Period(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Algeria</td>
<td>Large-scale civil war after 1991 military coup</td>
<td>1991-</td>
</tr>
<tr>
<td>2</td>
<td>Burundi</td>
<td>Ethnic civil war of the Hutu against the Tutsi-dominated government</td>
<td>1994-2005</td>
</tr>
<tr>
<td>3</td>
<td>Chad</td>
<td>Southern armed groups challenged the political and economic dominance of the north</td>
<td>1994-1998</td>
</tr>
<tr>
<td>4</td>
<td>DR Congo</td>
<td>Uprising against President Mobutu followed by transnational proxy wars</td>
<td>1994; 1996-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1998; 1994-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1999</td>
</tr>
<tr>
<td>5</td>
<td>Djibouti</td>
<td>Ethnic insurgency of the Afar to achieve political participation</td>
<td>1991-1994</td>
</tr>
<tr>
<td>6</td>
<td>Guinea-Bissau</td>
<td>Uprising against President Vieira to enforce a change in government</td>
<td>1998-1999</td>
</tr>
<tr>
<td>7</td>
<td>Ivory Coast</td>
<td>Irregular war after mutinous soldiers picked up arms against the government</td>
<td>2002-2004</td>
</tr>
<tr>
<td>8</td>
<td>Liberia</td>
<td>Irregular civil war aimed at removing President Taylor from power</td>
<td>2000-2003;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2001</td>
</tr>
<tr>
<td>9</td>
<td>Rwanda</td>
<td>Ethnic civil war after the Hutu-dominated government was removed from power in 1994</td>
<td>1994-2002</td>
</tr>
<tr>
<td>10</td>
<td>Sierra Leone</td>
<td>Irregular civil war against the state followed by infighting amongst rebel organizations</td>
<td>1991-1999</td>
</tr>
</tbody>
</table>

Table 5.3: Overview of the GED cases under investigation
Table 5.4: Civilian Casualties in the GED Dataset

<table>
<thead>
<tr>
<th>Distance to Capital (normalized)</th>
<th>Population (logged)</th>
<th>Line of Sight</th>
<th>Elevator</th>
<th>Military Casualties</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>0.04</td>
<td>0.06</td>
<td>0.08</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td>(11)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Significance levels: ∗∗∗ p < 0.001, ∗∗ p < 0.01, ∗ p < 0.05.
6 Civilian reactions to indiscriminate violence

The previous chapter tested the distance-decay effect of violence in insurgencies based on conflict events data. The causal logic in this setup is straightforward: Longer distances from the actors’ centers of operation have a negative effect on the actors’ ability to apply violence selectively. The corresponding empirical analysis is relatively simple: Static, geographic variables were coded for each conflict event to predict the type of violence applied by the actors. Unfortunately, systematic reactions to specific types of violence are harder to analyze quantitatively. In this chapter, I will discuss some of the methodological challenges of analyzing reactive patterns in events data and propose a solution. A Monte Carlo analysis is used to demonstrate the effectiveness of this solution in analyzing reactive patterns in conflict events data. Corresponding results are presented in the supplementary information (section 9.3 on page 171). Based on this new method, I analyze reactions to indiscriminate violence in Afghanistan in the second half of this chapter.

6.1 Introduction

The study of political violence has benefited in recent years from a rapid increase in the availability of conflict events datasets (Raleigh and Hegre, 2005, 

\[1\] This chapter and the corresponding supplementary information partially build on “Matched Wake Analysis: finding causal relationships in events data”, a paper prepared for presentation at the Annual Convention of the Swiss Political Science Association, Zurich, Switzerland, January 31–February 1, 2013, written in collaboration with Karsten Donnay.
Sundberg et al., 2011, Leetrau and Schrodt, 2013, SIGACT, 2010a,b). In these data, single instances of violence are coded together with their geographic coordinates and the date of occurrence. Several recent publications successfully explore some of the micro-dynamics of civil conflict by analyzing such data: Spatial determinants of violence in civil wars and the extent of primary conflict zones have been identified in a series of publications (Hegre et al., 2009, Raleigh and Hegre, 2009, Buhaug, 2010). The geography of ethnic settlement patterns has been linked to conflict onset and intensity (Weidmann, 2009, 2011). Research into the causes and effects of violence against civilians in civil wars (Kalyvas, 2006, Lyall, 2009, Kilcullen, 2010) and patterns of “tit-for-tat” dynamics between political actors (Jaeager and Paserman, 2006, 2008, Haushofer et al., 2010, Linke et al., 2012) has drawn on conflict events data. However, while progress has been made in relating conflict intensity to geographic conditions, more complex endogenous mechanisms that drive conflict at the micro-level remain largely elusive to quantitative analysis, despite their theoretical prominence. In principle, events data provide empirical means to study the changes in the trajectory of the conflict brought about by specific incidents. However, the very nature and strength of such data – detailed information on individual conflict events, including their type, timing and location – also poses a central methodological challenge that some studies fail to address: the absence of natural spatial and temporal units of analysis. Some studies have circumvented this problem by relying on artificial units of analysis, such as grid-cell-months, for aggregating event counts and coding covariates. While introducing these artificial units conveniently clears the way for econometric analysis, their arbitrary size is a source of random variation and bias (Openshaw and Taylor, 1979, Dark and Bram, 2007).

Closing the gap between rich event-level data and methods suitable for testing competing micro-hypotheses is a critical challenge. For this purpose, a novel technique for causal inference in disaggregated events data is introduced that combines two techniques for ensuring robust and clean causal inference: sliding spatio-temporal windows (Kulldorff, 1997, Schutte and Weid-
mann, 2011, Braithwaite and Johnson, 2012) and statistical matching (Rubin, 1973, DiPrete and Engelhardt, 2004, Iacus et al., 2012, Sekhon, 2011, Diamond and Sekhon, 2013). Beyond the realm of conflict research, this new method could be equally applied in quantitative criminology and epidemiology. Based on this method, the substantive question of whether indiscriminate violence really leads to increased support for the adversary will be treated in depth in the second half of this chapter. In doing so, the corresponding analysis draws on important insights into the micro-dynamics of civil war that have been generated recently.

Not surprisingly, two long-standing claims usually inform current thinking, as discussed in chapter 3. Deterrence-based explanations going back to Hobbes suggest that higher levels of violence against the insurgents might undermine their civilian support. Along these lines, civilians that witness the destructive capabilities of the state refrain from joining the uprising out of fear for their lives. At the same time, the existing rebel movement faces a more severe collective action problem due to increased risks for individual combatants, which in turn generates incentives for free riding (Olsen, 1965, Tullock, 1971, Lichbach, 1995). Another line of thinking puts a greater emphasis on the quality of applied violence. Indiscriminate violence is assumed to lead to more rebel mobilization due to civilians joining the rebel forces in retaliation for innocent bystanders that were harmed by the perpetrator. Avoiding reactive mobilization is an essential part of counterinsurgency doctrine. Most recently, a series of empirical studies have tried to shed light on the existence of reactive patterns in insurgencies to solve the long-standing dispute between advocates of deterrence and alienation (Downes, 2007, Kalyvas and Kocher, 2009, Lyall, 2009, Kocher et al., 2011, Condra and Shapiro, 2012, Linke et al., 2012, Braithwaite and Johnson, 2012). Interestingly, the empirical findings lend some support to both lines of thinking, leaving the discussion unresolved.

As discussed in chapter 3, I expect alienation to be the dominant effect of indiscriminate violence. However, it is difficult to derive an informed expectation about the spatial and temporal distances between instances of indiscriminate
violence and the reactions they cause. This chapter draws on a newly introduced method for an exploratory and non-parametric analysis of reactive behavior. Based on this methodology, a clear picture of reactive collaboration emerges which lends support to alienation-based accounts: Civilians increase their collaboration with the incumbent in response to indiscriminate insurgent violence. In response to indiscriminate incumbent violence, however, they decrease their support. This chapter therefore contributes to the theoretical understanding of reactive mobilization in insurgencies and introduces a novel empirical technique tailored to testing hypothesis 2. Beyond that, it is of direct political importance because it shows that deterrence in irregular war is a wrong-headed strategy. The existing literature will be reviewed in the next section. After that, the ongoing conflict in Afghanistan and the dataset used in the empirical analysis will be discussed.

6.2 Existing literature

The theoretical prominence of endogenous conflict dynamics has motivated a number of empirical studies that will be briefly discussed in this section. One important empirical distinction for conflict-event studies is whether or not they can draw on “natural” units of analysis. Some data are collected on the level of individual villages or hamlets. In such situations, the effects of specific types of events (“interventions” hereafter) on subsequent levels of violence can be estimated by comparing levels of violence before and after they occur using either statistical matching or multivariate regression analysis (Lyall, 2009, Kocher et al., 2011). In many comparable studies that draw on conflict events data, however, such natural units are absent. Conflict events that are precisely referenced to geographic coordinates and dates (Raleigh and Hegre, 2005, Sundberg et al., 2011, Leetrau and Schrodt, 2013, SIGACT, 2010a,b) are not tied to fixed units of analysis. Researchers therefore usually aggregate event counts within spatial units such as artificial grid cells and then analyze discrete spatio-temporal series using standard econometric techniques.
Unfortunately, there is usually no empirically or theoretically informed strategy for choosing the sizes of such cells.

Using artificial spatial units for statistical analysis leads to two problems widely described in the methodological literature. First, if cells of arbitrary sizes are the units of analysis, the number of available observations directly scales with the chosen cell size: the smaller the cells, the more observations. Of course, regular null hypothesis significant tests crucially depend on the number of available observations. As \( N \) increases, the standard errors usually decrease and even the smallest possible empirical signals become statistically significant. A second problem extensively described in the geographic literature is the “modifiable areal unit problem” (MAUP) — that is, the fact that the selection of artificial cell sizes drives various spatial statistics (Openshaw and Taylor, 1979, Openshaw, 1984, Dark and Bram, 2007). As eloquently put by (Openshaw, 1984, 3): “Quite simply, the areal units (zonal objects) used in many geographical studies are arbitrary, modifiable, and subject to the whims and fancies of whoever is doing, or did, the aggregating”. In many cases, this problem also pertains to spatial units that seem “natural” as they have existed for a long time by convention, such as postcode areas or administrative units. Yet such units may be arbitrary for certain types of empirical analysis. For example, when studying the geographic diffusion of crime in major cities, one might start by testing for spatial autocorrelation between spatial units, i.e. the fact that the dependent variable (crime rate) in one unit is correlated with the dependent variable in neighboring units at some point in time. In this case, it is immediately clear that the finding hinges on the level of spatial aggregation: When looking at city blocks in inner cities, positive spatial autocorrelation can probably be found, as criminals from one city block can easily reach the neighboring blocks. When choosing larger spatial units, such as postal code areas or electoral districts, the correlation might be less clear. Finally, when zooming out even further and analyzing crime rates for artificial spatial cells of, say, 20 by 20 kilometers, autocorrelation between inner cities and the surrounding suburbs is very likely to be small to non-existent. This effect comes about
because the actual diffusion process operates at small spatial scales while the analysis can be performed on arbitrary levels of aggregation.

Approaches to overcoming the MAUP have been proposed and applied in conflict research (O’Loughlin and Witmer, 2010, Schutte and Weidmann, 2011). A commonly used approach is “SaTScan” (Kulldorff, 1997), a method relying on sliding spatial and temporal windows to test for the existence of clusters of events on different levels of aggregation. Applied to epidemiology, SaTScan was originally introduced as a tool for testing whether a certain region faced an elevated per capita risk of disease. SaTScan analyzes clusters of point events in both space and time, using sliding spatio-temporal windows. The method allows for a fast assessment of whether event clusters could have been brought about by chance under corresponding distributional assumptions. To establish a baseline level of clustered events, SaTScan applies a simulation technique: For each size of the spatio-temporal window under consideration, the software allocates events randomly in space and time. Repeating this process in multiple iterations enables the analysis of a distribution of simulated events that was generated under baseline assumptions. Significant empirical deviations from this baseline can be identified for different cell sizes. Put differently, comparing this distribution of artificial events to the empirical record reveals how likely it is that the observed empirical patterns were brought about by chance.

In the epidemiological case of Kulldorff (1997), this baseline is well justified, simply assuming a constant per-capita rate of instances of non-infectious disease. In conflict settings, however, finding suitable baselines is usually much more difficult. Instances of insurgent violence, for example, are likely to result from a host of factors, including geographic exposure and reaction to previous violence. Randomly allocating events in space and time might not accurately capture plausible counterfactual scenarios: Instances of violence against civilians, for example, might be simulated to take place in uninhabited areas and a simulated baseline would not reflect the causal order of events found in the empirical record.
Relaxing the assumption of a uniform spatial distribution of events, Braithwaite and Johnson (2012) apply a permutation test within the framework of sliding spatio-temporal windows to the analysis of violent events in Iraq. In this setup, a random baseline is also simulated, but not by relocating conflict events in space and time. Instead, events remain in their original positions but event categories are randomly assigned. By holding constant the location and timing of events while changing event categories, a baseline scenario can be established in which event types are independent of one another. Comparing this simulated baseline to empirical distributions of event categories reveals whether or not specific classes of events tend to occur together, i.e. in clusters that are unlikely to have been brought about by chance. However, this measure of systematic co-occurrence does not establish a clear causal relationship between the event types.

In summary, a new quantitative methodology is needed that combines the robustness of sliding window designs with the inferential rigor of matching studies. This methodological contribution is all the more needed as endogenous conflict dynamics have received strong theoretical and increasing empirical attention. Clearly, Kalyvas (2006) has generated important insights into how and why violence is applied against civilians in civil wars. However, Kalyvas’ theory is mainly concerned with the deterrent and coercive effects of violence (Kalyvas and Kocher, 2007a, 210) and not how civilian alienation feeds back into the dynamics of conflict. With regard to the question of reactive mobilization, the theory assumes mobilization and civilian collaboration to be largely endogenous to military control (Kalyvas, 2006, 12; 118-132).

Two bodies of literature deviate from this assumption: Bootstrapping insurgent movements was an essential aim of communist insurgency, most notably to those inspired by Guevara’s foco theory (Guevara, 1961). The problem Guevara faced during the Cuban Revolution was that the country was not suited for revolutionary warfare according to Maoist doctrine, which put an emphasis on exploiting the vast areas of large countries when fighting the incumbent (Mao [1938] 1967, 7). Guevara broke with the assumptions of his
theoretical predecessor: "It is not always necessary to wait for all conditions favorable to revolution to be present; the insurrection itself can create them" (Guevara, 1961, 7). Creating favorable conditions for insurrection meant generating pockets of popular uprisings by mobilizing the civilian population.

Similarly, the counterinsurgency school has also focused on securing popular loyalties. High levels of violence are avoided as they tend to undermine the legitimacy of the attacker (Hastings, 2010, Ellsberg, 1970, Kilcullen, 2009, 2010). While revolutionary warfare and counterinsurgency approach the problem from slightly different angles, the underlying mechanism here is reactive mobilization: Instead of weakening the military opponent, violence can have the opposite effect. More civilians are alienated from the attacker and collaborate with the attacker's opponent.

In some sense the notion of reactive mobilization reverses the causal logic put forward by Kalyvas. Civilian collaboration is not so much assumed to be endogenous to military control, but military control can be brought about through mobilizing civilians. Clearly, the counterinsurgency school fought an uphill battle in conveying this effect to the higher echelons of the US Army in the 1960s. After all, classic metrics of military success rely on Loss Exchange Ratios (Biddle, 2006, 22), "body counts", or tonnage of dropped ordnance (Greiner, 2009, 23). A strong focus on which side could deploy more firepower in the field, inflict greater destruction on enemy cities, or endure a war of attrition for a longer period of time are core components of military deterrence.

Consequently, deterrence-based explanations continue to inform military and scholarly accounts. Deputy National Security Advisor Robert Blackwill suggested an aggressive air campaign against the Pashtun areas of Afghanistan: "[Under constant aerial attack] Taliban civil officials – like governors, mayors, judges and tax collectors – would wake up every morning not knowing if they would survive the day in their offices, while involved in daily activities or at home at night" (Blackwill, 2010). Discussing the prerequisites of successfully deterring insurgency, van Crefeld (2008, 235-246) presented the 1982 Hama massacre in Syria as a textbook case and derived rules for brutal counterin-
surgency: “There are situations in which it is necessary to resort to cruelty [...] The more like a thunderbolt out of clear sky it comes, the greater the effect. [...] once you have made up your mind to strike, you cannot strike hard enough” (p. 241). Drawing an analogy to his homeland, van Crefeld even advertised the counterfactual benefits of his hawkish rhetoric and applied it to the First Intifada of 1987: “[Yitzhak Rabin] could have ordered his troops to kill (say) five hundred Palestinians more or less on the spot and also blow up some object of high symbolic value. [...] the outcome could have been, if not peace, at any rate quiet” (van Crefeld, 2008, 242).

Clearly, this reasoning has lost much of its credibility in light of recent events in Syria. Bashar Assad’s heavy-handed reaction to an initially peaceful uprising in 2011 has lead to strong resistance. Despite a lack of outside support, former civilians and army defectors have resisted the state military for more than two years in a conflict that has claimed tens of thousands of lives.

To summarize the existing literature, almost opposite effects of violence on reactive mobilization have been proposed. Revolutionary warfare and counterinsurgency doctrine put an emphasis on the quality of applied force. Only selective and conditional violence is assumed to weaken the adversary while indiscriminate and unconditional violence have a positive effect on the opponents’ ability to mobilize. Deterrence-based accounts suggest that the quantity of applied force makes all the difference, with violence generally having a negative effect on the opponents’ ability to mobilize. Higher levels of force are assumed to exacerbate the collective action problem for the adversary. Finally, on this crucial point, Kalyvas (2006) – otherwise certainly a gold standard for theories on irregular war – is somewhat agnostic, assuming that civilian collaboration is largely endogenous to military control, while acknowledging the potential negative consequences of indiscriminate violence for the perpetrator (pp. 144, 150).

Most recently, several empirical studies have tried to resolve the theoretical debate. In a thorough analysis of reactive insurgent violence in response to the indiscriminate shelling of villages in Chechnya, Lyall (2009) reports a negative
effect of indiscriminate violence on insurgent activities, seemingly confirming deterrence-based explanations. The random artillery shelling of villages in the conflict zone led to a significantly reduced number of insurgent attacks in these villages in subsequent weeks and even months. In order to isolate the effect of indiscriminate shelling, Lyall (2009) applied statistical matching and also tested for spillover effects in terms of elevated levels of insurgent violence in neighboring villages after indiscriminate shelling took place. Downes (2007) contributed a case study from the Second Anglo-Boer War (1899-1902) and found that the size of the conflict zone affects to what extent indiscriminate violence undermines rebel support: the smaller the conflict zone, the stronger the deterrent effect.

However, empirical evidence in favor of alienation-based reasoning has also been published recently. Based on a high resolution GIS dataset that covers multiple years of the Vietnam War, Kocher et al. (2011) were able to show how military control systematically shifted in favor of the insurgency in response to largely indiscriminate aerial bombardments. Based on conflict events data from Iraq aggregated into administrative districts and weeks, Condra and Shapiro (2012) found a distinct reactive effect: Civilian casualties caused by incumbent forces lead to more subsequent insurgent violence, while civilian casualties brought about by insurgents lead to less insurgent violence.

Linke et al. (2012) applied a time-series analysis to finding reactive patterns between insurgent and incumbent violence in Iraq, drawing on a high-resolution conflict events dataset. Since these data are not aggregated to meaningful natural units of analysis, such as villages, the authors collapsed them into artificial cells of fixed sizes. They then calculated whether violent events predict violent reactions for different temporal lags.

Building on a similar approach, Braithwaite and Johnson (2012) analyzed the spatial and temporal clustering of conflict events in Iraq. Instead of reallocating conflict events spatially and temporally, the authors let events remain in their original position, but shuffled event labels to simulate random baselines against which they compared the empirical record. Similarly, several studies
on the location and diffusion of conflict events have applied scan statistics and simulated baselines (O’Loughlin and Witmer, 2010, Schutte and Weidmann, 2011). While these studies correctly address the MAUP, the simulation of baselines also introduces natural limitations to the inferential insights that can be obtained in this way. For example, Braithwaite and Johnson (2012) find certain event types tend to cluster more strongly than they would under the simulated independence assumption, but they cannot identify the causal effects of certain events on others with methodological rigor comparable to the matching designs of Lyall (2009) or Kocher et al. (2011).

To overcome this problem, I developed “Matched Wake Analysis” (MWA) in collaboration with Karsten Donnay (Schutte and Donnay, 2013). MWA draws on existing solutions for analyzing reactive patterns in events data, but combines the best of both worlds: sliding spatio-temporal windows to overcome the MAUP and an automated matching design that allows for clean causal inference. The next section will detail the methodological setup of an empirical analysis of the effects of indiscriminate violence in Afghanistan.

6.3 Matched wake analysis

Any attempt to overcome the discussed methodological shortcomings in the analysis of causal relationships in conflict events data must start with a theoretical understanding of the data generating process. A first crucial insight is that events come into existence through a variety of different mechanisms. In conflict research, there is the widely described effect of exogenous geographic conditions that drive overall levels of violence (McColl, 1969, Hegre et al., 2009, Raleigh and Hegre, 2009, O’Loughlin and Witmer, 2010). For example, locations that are more accessible to rebels than to the government, such as the mountainous tribal areas in eastern Afghanistan, see higher levels of violence. Ethnic settlement patterns have been linked to conflict events in Iraq (Weidmann and Salehyan, 2012). For conceptual clarity one can refer to them as the general exposure of any location to violence. Furthermore, levels of vio-
lence generally vary over time. A negotiated ceasefire, seasonal cycles, and the tactical initiative changing hands may all drive the intensity of conflict across a war zone. These aspects can be referred to as the *momentum* of a conflict at any given time.

![Diagram of conflict event types](image)

**Figure 6.1:** Concept of the “Matched Wake Analysis” in Afghanistan. In this sketch, stars represent instances civilian assistance to US forces. Classes of conflict events are categorized into “treatment” and “control” groups (here depicted as a triangle (a) and a rectangle (b)). A matched sample is generated based spatial covariates (*exposure*) and event counts and trends for different types of conflict events preceding the event under investigation (*momentum*). Finally, a Difference-in-Differences analysis allows for an estimation of the before-and-after average treatment effect (*reaction*).

### 6.3.1 Sliding window design

Isolating the effects of exposure and momentum on the number of dependent events is a crucial prerequisite for cleanly analyzing the causal effect of interventions. Figure 6.1 illustrates the logic of this empirical strategy.

In this conceptual sketch, three types of conflict events are depicted. The rectangular symbol in the center of the right cylinder represents an instance of violence assigned to the “control” category, i.e. an intervention to which the “treatment” category is compared. The triangle in the center of the left cylin-
der represents an instance of violence from the “treatment” category which is assumed to have a specific causal effect on the subsequent number of dependent events. Comparing the effects of treatment and control interventions directly can lead to false inferences, however. For example, the “treatment” event might occur in a region far away from the power center of the perpetrator, such as an airstrike that causes civilian casualties. While this event usually causes reactive violence in its proximity, there is simply no way to attack the perpetrator as it has no presence on the ground. Comparing the resulting non-increase to a “control” event, for example a direct encounter on the ground without civilian casualties much closer to the attacker’s power center, could lead to false inferences because reactive violence in this second case is a possibility while in the first case it is not. Instead of having established that “control” events lead to more reactive violence than “treatment” events, one would have simply compared types of violence under vastly different conditions.

To prevent this problem, geographic context information needs to be taken into account. In general, data expressing the overall exposure to violence for both intervention types must be coded. Spatial information on local elevation (Gesch et al., 1999), natural landcover Hansen et al. (2000), population figures (CIESIN, 2005), a measure of disaggregated wealth in the region (Nordhaus et al., 2006), average line-of-sight, distances to Kabul and Pakistan (Weidmann et al., 2010), the predominant ethnic group in the region (Pashtun or Hazara) (Wucherpfennig et al., 2011), as well as the season of the year in which the attack took place were calculated for each intervention event.

This choice of spatial confounding variables was driven by the following considerations: Geographic factors, such as accessibility and distances to power centers, population density, and wealth affect the quality of violence, as shown in chapter 5, and the ability of civilians to actively assist US forces: In remote areas firmly under insurgent control, active collaboration might be more dangerous than in areas close to the capital. To prevent the correlation between the probability of treatment assignment and reactions to treatment from driving the empirical results, the corresponding variables were included in the
matching procedure. The season was included because overall levels of violence undergo annual cycles as visible in figure 6.2 on page 119. Possibly, overall conflict intensity also affects civilian reactions to violence. The predominant ethnic group in the region was included because the US invasion of 2001 replaced a Pashtun-dominated government. The most intense fighting recorded in SIGACT also takes place in the Pashtun areas of southern and eastern Afghanistan. Hazara and Pashtun populations may react systematically differently to insurgent and incumbent violence.

Similarly, momentum of violence can be established for all conflict events by counting the number of previous dependent events. As figure 6.1 indicates, the lower half of the cylinder is subdivided into two halves to calculate a trend in the number of dependent events, which is flat in both cases depicted here (one conflict event in each of the first two quarters of the cylinders).

In principle, associating observations with static spatial covariates and dynamic counts of previous and posterior dependent events would be sufficient to generate a statistical sample for subsequent analysis. This setup, however, still does not account for the MAUP since the size of cylinders in space and time cannot be identified based on theoretical expectations: Why should events at a distance of 20 kilometers be counted while events at a distance of 30 kilometers are excluded? It is exactly this type of arbitrary coding that Openshaw and Taylor (1979) have shown to drive quantitative inference in undesired ways.

As pointed out in the previous section, solutions to this problem have been identified in terms of sliding spatio-temporal windows. In this setup, the entire procedure of counting previous and posterior events for every intervention is repeated for multiple sizes of spatio-temporal cylinders. This solves the problem of inference hinging on arbitrary cell sizes and allows us to distinguish between small- and large-scale effects empirically. For example, the effect of a treatment event on the level of dependent events might be stronger in its direct spatial and temporal vicinity and not affect more distant locations. The downside of this strategy is that for each observation in the statistical sample, the counting of previous and posterior dependent events has to be performed
for all spatial and temporal parameter combinations. For large numbers of observations, this can be computationally expensive.

6.3.2 Statistical matching

The previous section described how interventions are associated with counts of previous and subsequent dependent events for different spatio-temporal cylinder sizes. However, without explicitly accounting for confounding factors, causal inference in this setup still suffers from selection bias. In line with studies using natural spatial units of analysis (Lyall, 2009, Kocher et al., 2011) statistical matching is applied to compare treated and untreated observations under otherwise comparable conditions.

The general idea behind matching is to approximate as closely as possible experimental conditions in observational data (Rubin, 1973). In experimental settings, treatment is applied randomly and its effects are observed in comparison to an untreated control group. This type of randomization is critical for unbiased inference but frequently absent in observational data. To emulate randomization, several techniques have been proposed. In the most simple setting, a large quantity of observations for both treatment and control are available and exact matching can be applied. In exact matching, only those observations are retained in the treatment group for which a corresponding observation can be found in the control group with identical numerical values for all relevant confounding variables. This means that these observations only differ with regard to the treatment being applied or not. Clearly, under such ideal conditions, the treatment effect can be directly estimated through the difference in means between the groups for the dependent variable (Iacus et al., 2012, 1).

Unfortunately, such conditions are hard to find in practice. Usually the confounding factors between treatment and control observations are comparable, but not completely identical. Several strategies exist to alleviate this problem. One approach is to capture the effect of the confounding factors on the probability of treatment assignment in a propensity score model (Rosenbaum and
Propensity score matching essentially amounts to predicting the probability of treatment in a binary dependent variable regression model. The predicted probabilities of treatment assignment for each observation are used as the “propensity score” and observations from treatment and control group with similar scores are used in the subsequent analysis. Several procedures exist to quantify this score similarity based on nearest neighbor analysis or various distance metrics.

There is a practical problem associated with this technique for this particular application. A propensity score model requires as much care in post estimation as any other binary dependent variable model. Moreover, since the goal of matching is to increase balance, i.e. to make the empirical distributions of the covariates more similar, the balance must be assessed for each covariate before and after matching. In practice, researchers go back and forth between propensity score model specifications and assess the improvements in balance. Poorly performing propensity score models can very well decrease the overall balance and therefore completely defeat the purpose of matching.

Clearly, a more robust and unsupervised technique is needed for MWA: Due to the sliding window design, matching must be performed repeatedly for all spatial and temporal parameter combinations and manual readjustments after post-estimation analysis are not an option. A very recent and computationally efficient automated matching technique alleviates this problem: Coarsened Exact Matching (CEM) (Iacus et al., 2012). In CEM, substantially identical, but numerical slightly different values are collapsed into bins of variable sizes for each covariate. Matching is then performed for observations belonging to the same bins. Finally, a subsequent analysis is performed for matched observations, but with the original numerical values. CEM generates well-balanced datasets by choosing bin sizes for different variables based on their empirical distributions.²

²A possible alternative to CEM is Genetic Matching (GM) (Diamond and Sekhon, 2013). The problem with GM is that it is computationally less efficient. With matching having to be performed repeatedly in MWA, GM would have resulted in much longer run times.
6.3.3 Estimation of causal effects

There are several methods that are commonly used to estimate the causal effect of the treatment after matching is performed. For example, a Difference-in-Differences design (DD) (see Angrist and Pischke, 2009, 227-243) has been proposed and used in related empirical studies (Lyall, 2009). To assess the causal effect of the treatment, DD performs an OLS regression on the matched dataset to estimate changes in the number of dependent events brought about by the treatment. The dependent variable in this model is the number of dependent events after interventions. The number of dependent events before the intervention is also necessarily included in the model. Note that aggregating counts for each pre- and post-intervention period solves the problem of serial correlation that DD designs are otherwise prone to (see Bertrand et al., 2004, 252). Moreover, the setup accounts for changing conflict dynamics unrelated to the intervention by matching on the trend in the dependent variable before interventions. The resulting DD specification is then:

\[ n_{post} = \beta_0 + \beta_1 n_{pre} + \beta_2 treatment + u \]

In this model, the dependent variable, \( n_{post} \), is the number of dependent events after interventions, i.e. the number of events in the upper half of the cylinder in figure 6.1. The intercept, \( \beta_0 \), can be interpreted as the absolute difference in reactive events for the control group, i.e. the observations for which the binary treatment-variable is coded as \( 0 \). The estimate for the previous levels of “dependent” events in the lower half of the cylinder, \( \beta_1 \), is the estimated growth rate for both groups independent of treatment assignment. The quantity of interest in this model is the estimate for the binary treatment variable \( \beta_2 \). This value is the difference-in-differences caused by the treatment – or more precisely, the estimated within-subjects before and after treatment effect of the treated. Finally, \( u \) is the error term.

As discussed before, this setup will only yield reliable results if treatment and control events are selected in the sample that occurred under otherwise comparable conditions through matching. Detailed summary statistics for the
matching procedure in terms of the multivariate $L_1$ imbalance measure and
the percentage of common support (Iacus et al., 2012) are provided. $L_1$ is a
multivariate distance metric that expresses the dissimilarity of the distributions
of the covariates in treatment and control group. Distributions of confounding
factors are approximated in fine-grained histograms. Averaged normalized
differences between these histograms are expressed in the $L_1$ statistic ranging
from complete dissimilarity (1) to full congruence (0). A similarly intuitive
measure is common support, which expresses the overlap in values for the
confounding factors of the treatment and control group in percent (see Iacus
et al., 2012). 100% common support refers to a situation where the exact same
value ranges can be found for all confounding factors in both groups.

6.3.4 Limitations of the approach

While the underlying logic of matching observational data on confounding
factors is sound and widely used in empirical social science (see for example
Abadie and Imbens, 2006, Herron and Wand, 2007, Diprete and Engelhardt,
2004), spatio-temporal data introduces potential pitfalls. Most importantly,
the spatio-temporal cylinders around the interventions may overlap partially.
If they do, the Stable Unit Treatment Value Assumption (SUTVA) inherent to
matching is violated. This states that the treatment effect of any observation
is independent of the assignment of treatment to other units (Cox, 1958).
Violating this assumption can lead to biased estimates.

While SUTVA violations may indeed pose a problem to clean causal infer-
ence in this application, there are ways to mitigate this problem. First, spatio-
temporal overlaps are easily identified in empirical data. As described above,
counting previous and posterior instances of violence is part of data prepro-
cessing and multiple instances of overlapping treatment and control events can
be counted as well. The simplest way to avoid drawing false inference is there-
fore to check the data for overlaps of treatment and control events and select
subsets that are not affected by this problem. For example, a civil war might
go through phases of intense violence (e.g. summer offensives) and calmer pe-
periods. Researchers could test the causal effects of different types of events in the calmer periods to avoid false inference from overlapping events. However, empirical insights into the conflict dynamics would then be exclusively limited to such calmer periods instead of the entire conflict.

Second, if substantial numbers of overlapping cylinders cannot be avoided, data can still be analyzed. In this situation, the following problem must be accounted for: Interventions preceding the intervention under investigation can affect subsequent levels of dependent events. In this situation, the causal effect attributed to the intervention would in fact be the product of a specific mix of different interventions (a double treatment, for example). A simple remedy is to match the number of previous treatment and control events, i.e., to use them as confounding variables. This ensures that the interventions retained in the post-matching sample have similar histories of treatment and control events. Another effect of matching on previous interventions is that non-overlapping treatment and control events have a higher probability of being selected for the post-matching sample. This is due to the fact that overlapping cylinders tend to differ with regard to the previous number of treatment and control events for the simple reason that the earlier event will be counted as a previous event for the later one. This effect leads to a matched dataset with fewer overlapping events. A side-effect of this approach is that it decreases overall balance between the treatment and control groups with regard to exposure, because overlapping events yield similar values for the related spatial confounding factors.

A third alternative strategy is to simply remove overlapping observations from the sample. The obvious problem with this approach is the potential bias arising from non-random deletion itself. A benchmark analysis using simulated data is presented in the supplementary information for this chapter (see section 9.3 on page 171). It shows that this strategy still performed better than the baseline method, but appears to lead to less robust estimates for overlapping cylinders than matching on the number of previous treatment and control events. Based on these insights, the next section will quickly
recapitulate the theoretical expectations. In the empirical analysis, counts for previous treatment and control events are used as confounding variables, i.e. they are included in the matching procedure.

6.4 Theoretical expectations

Two theoretical approaches dominate the discussion on reactive mobilization: deterrence- and alienation-based explanations. In the empirical realm, both positive (Kocher et al., 2011) and negative effects (Lyall, 2009) of indiscriminate violence on insurgent activity have been reported. Deterrence-based explanations suggest a negative effect of indiscriminate violence on mobilization for the opponent. The mechanism at work is a collective action problem imposed on the opponent. If, for example, incumbent forces engage in acts of indiscriminate violence, sympathizers of the rebels are less inclined to join the uprising since it is assumed that they follow a risk-reward consideration with survival being their main goal. The corresponding hypothesis is therefore:

**H2.1:** Indiscriminate violence leads to more civilian cooperation with the perpetrator

Aside from deterrence-based reasoning, alienation-based accounts assume reactive collaboration with the military opponent to be a likely response to indiscriminate violence in insurgency. This assumes that the collective action problem can be solved through selective incentives: Whoever is affected by indiscriminate violence will attempt to take revenge on the perpetrator. Revenge, along these lines, qualifies as a selective incentive that potentially outweighs the perceived risks. This line of reasoning suggests the opposite effect:

**H2.2:** Indiscriminate violence leads to more civilian cooperation with the adversary

Instead of considering only those expectations that can be derived from the classic literature, other, more nuanced effects of indiscriminate violence should
also be considered. As mentioned before, neither of these two hypotheses has received decisive empirical support. The problem of making sense of and measuring reactive violence arguably stems from a problematic, simplifying assumption in the underlying theories. Both deterrence and alienation assume that civilians can change sides at a moment's notice, but this assumption is largely unrealistic. Civilians with little or no military training cannot strike back at the perpetrator immediately to avenge their fallen loved ones. Instead, they have to wait for a suitable opportunity to do so. Generally, opportunities for defection as a reaction to an incident arise only at some spatio-temporal distance from the incident, even though the wish to do so might be immediate. This line of thinking would suggest a more complicated picture of civilian collaboration and defection: In the direct vicinity of the trigger event, civilians are not inclined to collaborate with the adversary in fear of triggering additional violence by the perpetrator. Resentment against the user of force nevertheless builds up and ways to aid the adversary are silently considered. At some spatio-temporal distance, an opportunity for active assistance to the adversary and against the perpetrator arises. While still exposing themselves to as little risk as possible, civilians then take revenge. This effect makes an empirical analysis of reactive mobilization drastically more difficult, as the spatial and temporal lag between trigger and response are generally unknown. Nevertheless, the assumption that reactive behavior becomes visible only at some spatio-temporal distance from the incident needs to be considered. Compressed into a hypothesis, this expectation can be formulated thus:

**H2.3:** Indiscriminate violence only leads to more civilian cooperation with the adversary at greater spatio-temporal distances from the trigger event.

Clearly, types of micro-events in insurgency must be selected very carefully to test these hypotheses. Most importantly, a class of conflict events must be selected that allows for covert civilian assistance, i.e. an effective way to take part in the conflict at minimal risk. In the next section, I will discuss the dataset used for the empirical analysis.
6.5 Data

6.5.1 Temporal dynamics

In chapter 5, the spatial extent of the war in Afghanistan and the SIGACT were discussed. For the purpose of analyzing reactive dynamics, it is important to briefly review the temporal development of the uprising, which is rather typical for a large-scale insurgency: A small core of highly motivated combatants slowly attracts additional followers and mounts increasingly larger and more sophisticated attacks. Despite heavy losses, additional insurgents are mobilized and the conflict escalates over several years. This pattern in reflected in figure 6.2 which shows the conflict intensity over time. Besides the overall trend towards more insurgent activity, the figure also indicates annual variation in overall levels of violence. During spring and summer, activity generally picks up while the colder times of the year appear calmer. The plot on the left shows cumulative casualty counts. A simple deterrence logic would indicate that higher losses lead to less activity, but the reverse seems to be the case for Afghan insurgents. Reactive mobilization could partially explain these temporal dynamics.

6.5.2 Coverage

The empirical analysis requires highly accurate conflict events data. After having carefully considered alternative datasets, I decided to use SIGACT data coded by the US military in Afghanistan which was released to the general public by wikileaks.org on July 25, 2010.\(^3\) I decided to use this illegally distributed data in a responsible manner for basic research, given that the empirical analysis would not in any way harm or endanger the individuals, institutions, or political actors involved. To ensure this, my analysis only focuses on the events

\[^3\] The Hamlet Evaluation System (HES) is a Vietnam-era dataset that has been analyzed in comparable studies. The Armed Conflict Location Dataset (ACLED) covers a variety of conflicts, but does not distinguish among as many types of conflict events as SIGACT. The GED dataset puts a special emphasis on geo-referencing lethal events (Melander and Sundberg, 2011), but is also less explicit about which kinds of events led to casualties.
Figure 6.2: Development of the Afghan insurgency over time. The upper plot shows the number of actions per month for both incumbent and insurgent. The lower plot shows cumulative casualties for US Forces, insurgents, Afghan National Security Forces, and civilians. The so-called “fighting season”, an annual cycle in the intensity of the conflict that peaks in the summer, is clearly reflected in the data. Despite heavy casualties, insurgents were able to increase their activity over time.
in the statistical aggregate. Moreover, the particular empirical strategy of this study employs a matching design which entails that no marginal effects are estimated for confounding factors, which further strengthens the anonymity of the findings. Based on these precautions, the ethics committee of ETH Zurich reviewed a proposal for this study carefully and then allowed it to proceed.

As discussed before, the SIGACT ("Significant Activity") files cover the time period from 2004 to 2010 and amount to 76,247 records. All records are time- and geo-referenced, affiliated with insurgent or incumbent activity, and distinguish among 154 types of events. SIGACTs are passed up the chain of command from the platoon level, allowing for an extremely detailed record of the conflict from the US perspective. The potential limitations of SIGACT are discussed in chapter 5.

Several potential limitations arise with a sole focus on one side of the conflict. First, the figures for civilian casualties might be generally too low since it is the soldiers in the field who report them without independent confirmation. The use of indirect fire or air strikes in particular might harm bystanders without ground troops taking notice. Moreover, activities of other coalition troops, private contractors, and US service branches (such as the US Air Force) are not systematically recorded in the data. Insurgent-civilian relationships are, of course, also not visible in the data. Apart from these limitations, SIGACTs provide the most complete and arguably unbiased view on the Afghan War. Figure 6.3 illustrates the scope of the SIGACTs from Afghanistan and shows which of all the possible interactions between the actors are visible in the empirical record. Note that civilian relationships with the insurgents cannot be directly observed, but collaboration with US forces can be measured, as well as the type of violence used by the actors.
6.5.3 Eligible event categories

To ensure consistency across chapters, the event categories for selective and indiscriminate violence were the same as in chapter 5 (see section 5.1 on page 80): Selective incumbent violence was operationalized as direct fire engagements with insurgents. Indiscriminate incumbent violence was operationalized as both indirect fire and close air support. Similarly for the insurgent side, direct fire was counted as selective violence, while indirect fire and landmines were coded as indiscriminate violence. However, these categories do not provide information on civilian collaboration with the military actors.

In order to measure how loyalties change in response to indiscriminate violence, this analysis relies on a direct measure of civilian collaboration: the turning in of unexploded ordnance, or other explosive remnants of war that could be used by insurgents against US forces. Why should the turning in of unexploded ordnance be investigated in this analysis instead of other kinds of events?
To compensate for the lack of heavy weaponry, insurgents in Afghanistan often rely on “Improvised Explosive Devices” (IEDs) in attacks on both civilian and military targets. IEDs account for the largest number of US casualties in Afghanistan according to icasualties.org. In many cases, IEDs are military-grade explosives obtained from unexploded ordnance.

Due to this technical particularity, obtaining unexploded ordnance is a crucial prerequisite for generating a constant supply of new IEDs. Confronted with unexploded ordnance, civilians face a strategic choice: They can either remain passive and thereby allow explosives to be obtained by insurgents, or they can turn in explosive remnants of war. Taking sides with the incumbent would entail informing them of the threat at hand, thereby denying insurgents access to necessary supplies. Lives and material on the incumbent side could be spared due to such an action.

The other strategic option for the civilian population is to cooperate with the insurgents. A variety of possible implementations of such a cooperation spring to mind. Civilians could point insurgents and their collaborators directly to the explosives or just passively allow for ordnance to fall into insurgent hands. Either way, the turning in of material that aids the insurgency provides a rare opportunity to take sides in civil war without having to signal loyalties publicly, which would be extremely risky. In order to find out whether the proposed hypotheses hold, the SIGACT data for Afghanistan were analyzed based on MWA.

In order to identify reactive patterns in SIGACT, the potential interactions between the actors depicted in figure 6.3 must be operationalized. Moreover, the limitations inherent to data coded by military forces must be considered. Most importantly, the data do not hold direct observations of the interactions between civilians and rebels. Therefore, variation in the levels of civilian support to the incumbent was used to analyze the effects of violence, i.e. the “Collaborate/Defect” link in figure 6.3.
6.5.4 Results

Generally, the results show very clear reactive patterns. Indiscriminate incumbent violence decreases collaboration in direct comparison to selective incumbent violence under otherwise comparable conditions. For insurgent attacks, the reverse effect can be observed: Indiscriminate insurgent attacks increase collaboration with US forces in comparison with selective attacks.

Results are presented below in a series of contour plots, showing the $\beta$-estimates for the treatment term in the DD regression (see figure 6.4). The shaded areas in the plots indicate p-values above 0.1, i.e. parameter combinations for which no significant average treatment effect could be found. Dotted lines indicate p-values smaller than 0.1, but above 0.05. All non-shaded areas indicate significant effects for the interaction term. The “perspective” plot below is meant to give a more intuitive impression of the estimates for different spatio-temporal aggregations.

Indiscriminate incumbent violence affecting civilian collaboration

Civilian assistance to incumbent forces as a reaction to indiscriminate incumbent violence was operationalized as changes in the number of “turn in” events brought about by interventions. Most notably, indiscriminate incumbent violence has different effects for different spatio-temporal aggregations. Most locally, no strong effect can be seen in the analysis. However, as distances from the interventions increase up to 4 km and 20 days, a clear negative effect becomes visible. Note that this effect is not visible for the immediate vicinity of the attack site (<2 km), but if the spatial offset is taken into account, a robust negative effect emerges and remains visible along the entire time axis. An isolated positive effect that is weakly significant is also visible in one cell in the upper right corner of the plot. Since it is not robust to changes in spatio-temporal aggregations and not significant at the 95% level, it can be safely ignored. The results are clearly dominated by the negative estimate at medium distances from the interventions. How does this finding relate to the hypotheses? Clearly, hypothesis 2.1 and 2.2 would predict a homogeneous effect for
all parameter combinations. If deterrence was the mechanism at work (2.1), we would expect indiscriminate incumbent violence to lead to more collaboration with the state, i.e. positive estimates for all parameter combinations. Alienation (2.2) as an alternative mechanism would lead to negative estimates, reflecting declining collaboration with US forces in reaction to indiscriminate incumbent violence. But the empirical picture is more complicated and best described by hypothesis 2.3: Declining collaborations with the perpetrator of indiscriminate violence and increased collaboration with the adversary only become visible at certain spatio-temporal distances from the trigger event. However, alienation is the predominant effect of indiscriminate violence according to these results. It is important to note, however, that the effect is very small: At 20 days and 4 km, it is only -0.04. This means that for every 100 instances of indiscriminate violence, 4 fewer instances of civilian assistance to US forces occur.
Figure 6.4: Contour plot showing the before-and-after average treatment effect from the DD regression in the upper plot. The control group consists of instances of selective violence by US forces and the treatment group of indiscriminate violence by US forces. The dependent variable is the number of events in which civilians turned in unexploded ordnance. In the shaded areas, the estimate was not significant at the 95% level. Note that the estimates are predominantly negative. The lower plot gives a three-dimensional representation of the estimates.
Apart from these substantive findings, the effectiveness of the matching procedure was also assessed. Figure 6.5 shows summary statistics for each parameter combination in the MWA analysis. In the upper table, the columns in each cell indicate changes brought about by the matching procedure. The situation prior to matching ("pre") and after matching ("post") is described by a number of summary statistics. "T" and "C" stand for "treatment" and "control" and indicate the numbers of observations available for each group. "L1" and "%" show the two multivariate similarity measures introduced in section 6.3.3 on page 113. Both statistics compare the joint distributions of the confounding between treatment and control group, either with regard to the normalized dissimilarity between the distributions of confounding factors (L1), or the overlap of value ranges in the treatment and control groups (%).

The range of the substantively interpreted effect at distances of up to four kilometers and between 20 and 30 days sees a strong improvement in balance as a result of the matching. Common support is increased from about 10% to almost 40% in these cells. Similarly, L1 decreases from 0.84 to about 0.54. Across all spatial and temporal parameters, matching decreases L1 and increases common support, highlighting the effectiveness of coarsened exact matching.

The lower plot in figure 6.5 shows counts of treatment and control events before and after interventions, i.e. violations of the stable unit treatment value assumption. Instances of double treatment and treatment-control combinations are referred to as "double" and "spill", respectively. The graph reports overlaps prior ("<") and subsequent (">") to interventions, as well as overlaps affecting the whole cylinder ("<>”). For the substantively interpreted area, instances of temporally and spatially overlapping treatment interventions occur: For distances up to 4 km and 20 days, 38% of all observations in the sample have at least one preceding intervention of the same type within their spatio-temporal cylinder. As discussed before and tested in the Monte Carlo analysis in the supplementary information, a possible remedy in this situation is matching on the counts of previous interventions. Based on this precaution,
the overlapping interventions are unlikely to drive the statistical results.\textsuperscript{4}

\textbf{Indiscriminate insurgent violence affecting civilian collaboration}

In order to establish reactive collaboration with the adversary as a symmetric effect in insurgencies, the effects of insurgent violence on civilian assistance to US forces was also analyzed. As visible in figure 6.6, in ranges up to 4 kilometers and for temporal lags of 20 days and beyond, a clear reactive effect is visible. All significant estimates in this MWA analysis are positive, suggesting that indiscriminate insurgent violence leads to more collaboration with US forces. The spatial and temporal lag of the reaction corresponds well to the previous analysis of reactions to incumbent violence. A second effect is visible for distances up to 10 km and 30 days, but the figure is clearly dominated by the effect closer to the site of the intervention. Again, these insights correspond very well to hypothesis 2.3: Alienation from the user of force occurs, but only at certain spatial and temporal distances from the trigger event.

For this analysis, summary statistics were also calculated for the matching procedure and SUTVA violations were assessed. As visible on the left in figure 6.7, both the treatment as well as the control groups in this sample are much larger than in the analysis of reactions to incumbent actions. In the area of the plot where the treatment effect in the DD regression is significant, only about half of all control events are retained in the matched sample. A higher number of treatment events is selected into the matched sample, however. Unfortunately, the increase in common support and the decrease in $L1$ is less pronounced than in the previous analysis. The maximum common support in the post-matching sample is only 17.2\% for 20 days and 2 kilometers. Less than 10\% increase in common support due to matching can be seen in interpretable area of the plot. With regard to the possible SUTVA violations, the analysis also faces more severe problems. Instances of double interventions in the interpretable area of the plot are quite frequent, ranging from 0.5 at 2

\textsuperscript{4}In the supplementary information to this chapter (section 9.3 on page 171), robustness tests in terms of alternative confounding factors and placebo tests are also presented.
Figure 6.5: Summary statistics for the matching procedure can be seen in the upper table and counted SUTVA violations are indicated below. Note the substantial increase in the common support (%) and the reduction in the $L1$ distance metric brought about by the matching. Also note the substantial fraction of SUTVA violations, especially “double treatments” within the interpreted area. To account for those, the number of previous treatment and control events was included in the matching procedure.
Figure 6.6: Contour plot showing the before-and-after average treatment effect from the DD regression in the upper plot. The control group consists of instances of selective violence by insurgents and the treatment group of instances of indiscriminate violence by insurgents. The dependent variable is the number of events in which civilians turned in unexploded ordnance. In the shaded areas, the estimate was not significant at the 95% level. Note that the estimates are predominantly positive in the interpretable areas. The lower plot gives a three-dimensional representation of the estimates.
days and 20 km to 0.7 for 4 days and 50 km. Again, matching the exact numbers of previous interventions was performed to prevent the effects of SUTVA violations from driving the results.

6.6 Discussion

Rapidly growing interest in disaggregated conflict-event studies underlines the need for better methodology for causal analysis. Standard econometric approaches only work reliably if data are available in natural spatial units of analysis. In many scenarios, such data are absent and relying on artificial units bears the risk of generating false inference. Sliding windows in space and time have been previously applied in these contexts. While they adequately account for the MAUP, corresponding studies are rather weak on the inferential side: Usually, sliding window designs only demonstrate that spatial and temporal clustering in empirical data significantly deviates from the clustering that is expected under simulated baseline conditions.

Combining the best of both worlds, the presented method applies a sliding window and an unsupervised matching technique, offering an analysis of the causal connections between different types of events for different spatial and temporal distances from a given intervention. A caveat of this approach is that it requires the spatial or temporal separation of different types of interventions to cleanly attribute reactive effects. Such quasi-experimental conditions are not always present in empirical data. Nevertheless, assessments of the robustness of the method have shown that substantive inference can still be performed for situations of small spatio-temporal overlaps of interventions. Higher levels of overlaps can also be analyzed if numbers of previous treatment and control events are used as confounding factors in the matching procedure. In summary, a suitable tool to test the second main hypothesis of this study – reactive support for the adversary in response to indiscriminate violence – has been crafted and tested.

With regard to this substantive effect, deterrence- and alienation-based rea-
Figure 6.7: Summary statistics for the matching procedure on top and counted SUTVA violations below. Note the increase in the common support (%) and the reduction in the $L1$ distance metric brought about by the matching. Also note the substantial fraction of SUTVA violations, especially “double treatments” in the lower table. To account for those, the number of previous treatment and control events was included in the matching procedure.
soning has dominated the discussions on the effects of violence in civil wars for decades. Recently, a series of mixed empirical results have given support to both camps. The results presented in this study clearly indicate the existence of alienating effects in the Afghan insurgency, but they also highlight the importance of accounting for different spatial and temporal offsets in the analysis of reactive events. While the signs of the estimate are consistent across spatio-temporal windows, the exact estimate varies. Reaction to indiscriminate violence only becomes visible at certain spatial and temporal offsets. The corresponding lag might result from risk aversion in civilian behavior, or from the fact that civilians that seek to collaborate with the incumbent have to wait for an opportunity to do so. Substantively, the results strongly support alienation from the perpetrator as the predominant reaction to indiscriminate violence. The unprecedented detail and wealth of information encoded in SIGACT allows us to establish this finding based on a direct measure of civilian support. The results are somewhat in line with the findings of Condra and Shapiro (2012) for Iraq and Kocher et al. (2011) for Vietnam.

This finding is in line with a basic insight of the counterinsurgency school: Loyalties shift towards the strategic adversary if actors cause indiscriminate destruction in the field. If actors react to increased resistance by applying more force of the same kind, they can quickly find themselves locked in an increasingly deadly struggle in which tactical victories over enemy combatants translate to strategic defeat in population-centric warfare. This mechanism could explain how peripheral insurgencies without strong international support can take on mechanized armies, as is happening in Syria, for example. Historical examples of insurgencies in developing countries that have defeated superpowers also seem more easily explainable based on this mechanism. The crucial policy implication here is that the logic of breaking enemy resistance or inducing fear through overwhelming force seem counterproductive and dangerous in light of the presented evidence.

With regard to the wider scope of this study, the empirical reality of the second proposed mechanism - reactive support for the adversary - has been es-
tablished, at least for one typical case of insurgency. The remaining challenge is therefore to show that the presented theory generates suitable predictions for the macro-level. Numerical predictions have been generated with the simulation model in chapter 4 and suggest that a fundamental property of the human geography of different countries should have a profound effect on insurgencies. This territorial balance essentially expresses which of the military actors has the means of exercising selective violence over the bulk of the population. The following chapter will describe an empirical operationalization of this territorial balance and test whether the simulated predictions correspond to the estimated effect in empirical data.
7 Territorial balance: 
geography, outcomes, and 
casualties

7.1 Introduction

This study set out to articulate an integrated theory of insurgency. The previous chapters have provided insights into how variation in the quality of applied violence and reactive mobilization shape these types of conflicts: Violence becomes more indiscriminate as the distance to the power center of the perpetrator increases. Moreover, indiscriminate violence leads to increasing public support for the adversary, as demonstrated in the previous chapter. In addition to having been subjected to empirical tests, these assumptions were used to simulate artificial insurgencies in an agent-based model in chapter 4. With the help of the simulation model, numerical predictions for the most probable military outcome and the approximate level of casualties have been generated under varying geographic conditions. In other words, the simulation model has generated numerical expectations for the macro-level properties of insurgencies (outcomes and casualties) under the assumption that specific micro-mechanisms ("Loss of Accuracy" and "reactive mobilization") shape these aggregate properties. In this final empirical chapter, the simulated results and the corresponding hypotheses will be compared to the predictions of outcomes and casualties of econometric models, which were fitted on a global sample of insurgencies that took place between 1970 and 2010.
To recapitulate, the proposed theoretical mechanism that connects geography to outcomes and casualties is the following: The military actor that is more remote from the civilian population resorts to higher levels of violence against civilians. The resulting backlash in popular support means that the other side receives stronger public support and mobilizes at a faster rate. This effect translates to a higher chance of military victory for the actor closer to the population. Moreover, if both actors are remote from the population, repeated cycles of indiscriminate violence and reactive mobilization drive the casualty figures.

In the agent-based model, geographic conditions that affect the actor’s abilities to apply violence accurately was operationalized as intersecting “Loss of Accuracy Gradients” (LAGs) that could be controlled by a single parameter, the territorial balance (TB). TB effectively defines the size of areas of the simulated space in which the actors can attack enemy combatants with superior accuracy. High parameter settings favor the incumbent side, while lower settings give the insurgent side an advantage. Over a large number of simulations and a wide parameter range, the TB parameter drives the outcomes of simulated insurgencies. More specifically, TB has a negative effect on insurgent success under the Lockean assumption that indiscriminate violence leads to reactive mobilization, but a positive effect on insurgent success for Hobbesian settings. Both configurations predict an inverse U-shaped effect of territorial balance on casualties.

In order to test these predictions, an empirical indicator that corresponds to the territorial balance in the simulation model is defined in this chapter and used to predict outcomes and casualties. As argued in chapter 5, I assume the incumbent power center to be the capital city, while the insurgent power center is the periphery. Based on geo-coded data on the global distribution of population (CIESIN, 2005), a measurement expressing population distances from the capital can be defined.

In a subsequent inferential analysis, I show that the corresponding indicator is significantly related to the possible outcomes of insurgencies: Countries
with centrally concentrated populations are more likely to defeat peripheral insurgent movements, and countries with population concentrations in the periphery are more likely to see insurgencies succeed in toppling the government. Moreover, medium levels of this indicator are associated with higher casualties in insurgencies, while high and low values correlate with comparatively bloodless struggles. These results further corroborate the presented theory as they match the functional forms of the simulated Lockean predictions.

7.2 Existing literature

While civil war onset is the most prominent dependent variable in contemporary civil war studies (for example Fearon and Laitin, 2003, Buhaug et al., 2008, Cederman et al., 2010, Ward et al., 2010, Deiwiks et al., 2012), a series of studies have attempted to explain the duration, severity, and outcome of civil wars based on conflict-level characteristics (see Collier et al., 2004, Hegre, 2004, DeRouen and Sobek, 2004, Lacina, 2006, Cunningham et al., 2009, Buhaug et al., 2009, Wucherpfennig et al., 2012). Mack (1975) points to the asymmetry in incentives for third-party actors and domestic insurgents as an explanation for why some insurgencies prevail against all material odds: For insurgent movements, winning the war is generally a matter of survival, while external powers can withdraw. Moreover, foreign occupation creates national cohesion for the insurgents, while the domestic audiences of external actors are divided over the costs and benefits associated with the conflict. Goodwin (2001, 106) argues that the fate of “revolutionary movements” in southeast Asia can be best explained as a legacy of different types of colonial rule: Moderate policies towards nationalist movements prevented the communist opposition from harnessing nationalist sentiments for their own ends. While theoretically compelling, this mechanism neglects the conflict dynamics that affect violence and mobilization after the onset of open hostilities. Focusing his own study on how violence is used strategically in civil wars, Kalyvas (2006, 3) notes that “almost every macrohistorical account of civil wars points to the importance
of preexisting popular allegiances for the war’s outcome, yet almost every microhistorical account points to a host of endogenous mechanisms”.

Unraveling such endogenous dynamics, Arreguin-Toft (2001) analyzes the strategic interaction of the military actors, arguing that increased mechanization on the incumbent side and more reliance on guerrilla warfare on the insurgent side have been the dominating trends in military doctrine since World War II. Due to this specific interaction, insurgencies have been able to succeed at a higher rate than before. A variant of this argument is that modern warfare prevents incumbent forces from forging ties with the civilian population that are vital in counterinsurgencies (see Lyall and Wilson, 2009).

Focusing more on actor-level attributes, Cunningham et al. (2009) analyze the duration and outcome of civil wars from a dyadic standpoint instead of relying on country-level indicators as proxies for attributes of the actors. They find that strong rebel organizations lead to shorter civil conflicts. Similarly, rebel movement that have a non-violent political wing are often associated with shorter conflict durations. These results suggest that the possibility for achieving strategic objectives without having to rely on political violence has a pacifying effect on ongoing conflicts. Finally, the spatial constellation of military actors as well as geographic conditions in the conflict zone have received increasing attention. Buhaug et al. (2009) analyze the duration of civil conflicts and find that rebels that operate from the periphery contribute to longer conflicts, due to the central governments’ inability to control such remote territories. In a lengthy study directly tailored to explaining the outcomes of insurgencies, Connable and Libicki (2010) found that two variables are robustly related to the military outcomes of insurgencies: External support for insurgents generally increases their chances of success, while the use of terrorist tactics by insurgents has a negative effect on the probability of insurgent victory. This quantitative finding mirrors an earlier emphasis on the role of external support by Record (2007). With the exception of Buhaug et al. (2009), the existing literature has not incorporated the spatial constellation of military power centers. However, other relevant variables, such as external
support of the rebels and state strength have been identified. Beyond demonstrating the effect of their main independent variable – modern combat tactics – on insurgency outcome, Lyall and Wilson (2009) also share a comprehensive data collection on two centuries of insurgencies with the research community. Drawing on their dataset outcomes of insurgencies can be researched in a cumulative fashion (see Hegre and Sambanis, 2006). New explanatory variables can be tested against established predictors. Clearly, a prerequisite for testing my theory on the conflict level is a suitable method for operationalizing the spatial constellation of the military actors.

7.3 Theoretical expectations

In order to measure the proximity of the actors to the bulk of the population, assumptions about the locations of these power centers must be made. As in chapter 5, I equate the center of state power with the capital city and the rebels’ realm with the remote periphery. A suitable proxy for the government’s ability to control the civilian population with accurate force is therefore the capital’s distance to the bulk of the population. Similarly, the rebels’ ability to apply force accurately over large portions of the population relies on civilians being concentrated in the periphery. Based on the results of chapter 5, the more remote actor is more likely to use violence indiscriminately. Under the Lockean assumption that indiscriminate violence leads to reactive mobilization which is corroborated by chapter 6, the following empirical expectation can be formulated:

\textbf{H3.1a}: Population concentrations close to the capital city have a negative effect on the probability of insurgent victory.

However, the simulation model has also generated predictions for situations in which indiscriminate violence has a deterrent effect. Under such conditions, the effect of territorial balance is reversed. Although chapter 6 did not produce corresponding evidence on the micro-level, the theoretical prominence of
deterrence-based reasoning justifies the inclusion of a corresponding hypothesis:

**H3.1b:** Population concentrations close to the capital city have a positive effect on the probability of insurgent victory.

With regard to the effect of territorial balance on casualties, the theoretical predictions of Hobbesian and Lockeian accounts converge. Based on the presented theory and findings in previous chapters, the territorial balance determines the actors' ability to apply force accurately and thereby affects the number of casualties in insurgencies: If one side enjoys far superior accuracy due to proximity to the bulk of the population, the indiscriminate tactics of its opponent quickly undermine civilian support, leading to a comparatively bloodless struggle. However, if the bulk of the civilian population is located at medium distances to both power centers, it suffers from indiscriminate attacks by both sides.

The resulting reactive mobilization feeds into the circle of violence since it only results in the application of indiscriminate force. Multiple iterations of violence, alienation, and mobilization generate higher levels of deaths and destruction. The main hypothesis regarding casualties therefore expresses this inverse U-shaped relationship:

**H3.2:** Distance between population concentrations and the capital has a negative quadratic effect on the number of war casualties.

### 7.4 Empirical analysis

Drawing on the dataset by Lyall and Wilson (2009) that provides information on the outcome of insurgencies as well as data on casualties from the Correlates of War Project (Sarkees and Wayman, 2010) and the CSCW/PRIO battle death data (Lacina and Gleditsch, 2005), the proposed theory is tested in a large-N conflict-level setting. I operationalize the actors' proximity to the bulk of the population for the country level, making this study compatible
with other quantitative studies of insurgency (Fearon and Laitin, 2003, Lyall and Wilson, 2009, Connable and Libicki, 2010). A conflict-level measurement, the territorial balance indicator (TBI), will be introduced below, followed by the empirical results and a discussion of their implications.

7.4.1 Measuring territorial balance

In order to test the introduced hypotheses, the distance between the bulk of the population and the military actors must be formally defined. Data on the spatial distribution of population – the Gridded Population of the World Dataset (GPW) (CIESIN, 2005) – and the location of capital cities and international boundaries (Weidmann et al., 2010) were combined to form a single indicator for “territorial balance”. Intuitively speaking, the territorial balance indicator measures the fraction of the total population located within a certain distance from the capital city. For any normalized distance \( D \) from the capital to the most remote inhabited spot, the fraction of the population \( p_{d<D} \) within that distance can be calculated and plotted, according to \( t_{b_D} = \frac{\sum p_{d<D}}{\sum p} \), where \( p \) is the total population of a country. This measurement naturally falls into the interval \([0,1]\). With growing distance, \( t_{b_D} \) values form a curve, eventually connecting the points \((0,0)\) and \((1,1)\). Figure 7.1 shows the resulting curve for the United States, based on the 1990 spatial distribution of the population.

How can we interpret the exact shape of the curve in figure 7.1 on page 143? The curve simply expresses the fraction of the population that can be reached for any given distance from Washington DC. At first, the curve climbs rapidly as the densely populated east coast is covered. Around a cumulative population value of 0.4, the curve flattens out as the less densely populated Midwest is reached. It flattens out even further as the Rocky Mountains are integrated and then suddenly spikes around a distance value of 0.7 caused by the densely populated west coast. The sparsely populated wilderness of Alaska and the vast Pacific Ocean between Hawaii and the continent appear as an almost flat section on the right end of the curve.

For the statistical analysis, the described indicator must be collapsed into a
single number for which the area under the curve (AUC) was chosen. Formally: \[ TBI = \frac{a}{a+b}, \] where \( a \) is the area under the curve and \( b \) the area above it. A caveat of this approach is that global data on population numbers only dates back to 1990. Therefore, the post-WWII period that is frequently the focus of civil war research cannot be analyzed entirely based on this indicator and the analysis is temporally restricted to the time period between 1970 and 2010.\footnote{This measurement is of course not the only possible operationalization of access to the civilian population for the military actors. As a robustness check, I coded an alternative version of the indicator using distances to the nearest major city. The cities coded in this dataset (Nelson, 2008) had at least 50,000 inhabitants in the year 2000, and most of them also existed back in 1970. The results are reported in the supplementary information (section 9.4) and are substantively identical to the ones presented in this chapter.}

7.4.2 Analyzing outcomes

The possible outcomes in insurgencies can be categorized on a spectrum ranging from victory to conclusive defeat for either side. To preserve the ordinal information in the dependent variable, I coded incumbent victory as 1, settlement of the hostilities without decisive victory of either side as 2, and insurgent victory as 3. Ordered Logistic regression models were used to assess the relevance of the proposed variables. The time period under investigation is 1970 to 2010. The reason for limiting the time period to this range lies in the static nature of the main independent variable. The TBI in its current form relies on geo-referenced data that can only be obtained from 1990 onwards. Since the aggregate TBI measure is not subject to drastic annual changes, \( 1990 \pm 20 \) years was chosen as a suitable time frame for the analysis which uses static 1990 TBI values as the main independent variable.\footnote{To rule out endogeneity, i.e. systematic effects of insurgencies on the TBI, I compared TBIs calculated based on 1990 and 2010 GPW data for conflict and non-conflict cases. For countries that experienced conflict between 1990 and 2010, the average change in TBI was -0.0014, while non-conflict countries changed with a slight increase to 0.0041. These changes are roughly equivalent to 0.1 to 0.2 standard deviations of the entire 1990 TBI sample. In both cases, no systematic differences in the TBI distributions were found in Kolmogorov-Smirnov tests. Visual comparisons of the TBI distributions and side-by-side comparisons of the TB curves for conflict countries can be found in the supplementary information. In order to rule out multicollinearity in the data, I calculated correlations between the main independent variables. The resulting table can be found in the supplementary material and shows no signs of strong multicollinearity.}

For countries whose international boundaries have changed over time, I calculated the TBI...
Figure 7.1: TBI curve for the United States. The curve describes what fraction of the US population is within a certain distance from Washington D.C. The TBI expresses the area under the curve divided by the whole area of the plot. With the United States' population balance in the East, the resulting TBI is comparatively high. A visualization of the GPW dataset (CIESIN, 2005) for the region is visible in the upper half of the plot.
for the boundaries that were in place when the conflict broke out. Mounting methodological objections against a sole reliance on p-values (Ward et al., 2010, Schrodt, 2010) and the aspirations of this study to contribute to generating forecasts for current and future conflicts require additional assessments of the predictive capabilities of the models.

Both in-sample and out-of-sample, predictive performance was tested and corresponding results are included in the empirical analysis. To create a baseline for the statistical analysis, model 1 uses the model from the Lyall and Wilson (2009) study. The main independent variable of their study – modern combat tactics – had to be omitted in this analysis. Modern combat tactics according to Biddle’s (2006) definition were generally introduced after WWI. Therefore, there is simply no variance on this indicator for post-1970 insurgencies. However, Lyall and Wilson (2009) still provide a formidable point of departure both theoretically and empirically.

Model 2 only uses distance from the capital city of the incumbent to the conflict country and the TBI. The distance variable is equal to 1 in all cases where incumbents fight domestic insurgencies, but varies for out-of-area conflicts. Model 3 includes all introduced variables. Model 4 omits the distance variable and shows that the results do not change substantively in comparison to the full model. Model 5 was chosen to minimize the AIC value, a statistic that rewards high goodness-of-fit while penalizing for the number of explanatory variables.³

**Control variables**

Military and socio-economic control variables were obtained from Lyall and Wilson (2009) and supplemented with the main independent variable. As visible at first glance in table 7.1, several control variables are significantly associated with the ordinal coding of rebel success.

³The central result of the analysis holds for a wide range of model specifications. However, it was found that even the base model for the post-1970 period leads to quasi-complete separation: Extremely high values of the DISTANCE are consistently associated with rebel victory. Omitting the DISTANCE variable leads to essentially the same estimates for the remaining covariates, as model 4 shows.
Regime type one year prior to conflict onset \((\text{REGIME})\) and outside support to the rebels during the conflict \((\text{SUPPORT})\) are significantly and positively associated with rebel victory.\(^4\) The positive sign for the \text{REGIME} variable confirms the widespread suspicion that democratic governments are likely to give up fighting counterinsurgencies earlier than autocracies (see Lyall and Wilson, 2009). Domestic peace movements, for example, can put significant pressure on the incumbent in liberal and democratic societies. Autocratic systems, on the other hand, can keep fighting without having to take into account the interests of a broad electorate. Therefore, autocratic societies are more likely to defeat insurgent forces militarily. It should be mentioned, however, that this finding seems to be driven by the restricted sample size and does not hold for the entire post-WWI period, as reported by Lyall and Wilson (2009).

Foreign \text{SUPPORT} to the rebels is also reliably and positively associated with rebel victory. Foreign fighters willing to join the uprising, material support, and cooperating intelligence services can obviously strengthen the rebel forces, as proposed by Record (2007) and independently quantitatively confirmed by Connable and Libicki (2010).

The state capability indicator \text{POWER}\(^5\) is negatively and significantly associated with insurgent victory, which again confirms the intuition that stronger states are better at fighting off domestic challengers. The logged energy use of the incumbent as a share of the energy use of the whole population \((\text{ENERGY})\) also has a negative estimate, but fails to reach statistical significance. Despite the limited sample and the ordinal coding of the dependent variable, these results are largely in line with the results reported by Lyall and Wilson (2009).

An indicator of whether a country is occupied \((\text{OCCUPATION})\) is significant

\(^4\) \text{REGIME} was measured by the polity2 variable in the Polity IV dataset (Lyall and Wilson, 2009, Jaggers and Gurr, 1995).

\(^5\) \text{POWER} was coded as the natural log of the cumulative national capabilities of the incumbent in the last prewar year based on the Correlates of War dataset (v.3.02) (Lyall and Wilson, 2009, Singer and Small, 1994). It is important to mention that the variables \text{REGIME}, \text{POWER}, and \text{ENERGY} reflect the country's situation in the last prewar year, while \text{SUPPORT}, \text{HELI}, and \text{COLDWAR} code whether external rebel support, helicopter deployment, or parts of the Cold War took place at some point during the conflict.
in models 1 and 4, but not in the remaining models. Mean elevation of the
country (ELEVATION) does not reach significance at the 5% level. Clearly,
elevation averaged over the whole country is a rather crude measurement and
one should not read too much into this non-effect. The role of inaccessible ter-
rain and limited access to peripheral regions is very central in the theoretical
and qualitative literature and probably not adequately operationalized in this
variable.

A reliable indicator of war outcome is COLDWAR. In all four models,
COLDWAR is negatively associated with insurgent victory, implying that the
probability of insurgent success was significantly lower during the Cold War.
African countries that received military aid from the superpowers drive this
result: The Democratic Republic of the Congo, Liberia, and Rwanda have all
witnessed successful uprisings since the end of the Cold War. The variables
MECH and HELI reflect military capabilities of the incumbent. MECH en-
codes the degree of mechanization of the armed forces and has no significant
effect, which deviates from the findings reported in Lyall and Wilson (2009)
for the entire post-WWII period. The availability of helicopters (HELI) has
a positive effect on insurgent victory in all models. The number of languages
spoken in the conflict country (LANG) has a positive estimate in all models,
but is only weakly significant in models 4 and 5. This positive estimate is in line
with the discussion on cognitive access to the civilian population: Incumbents
should face a more profound problem of identifying civilian loyalties if multiple
languages are spoken in the conflict-torn country. International TRADE as a
fraction of the overall GDP is weakly negatively associated with outcome, as is
the natural log of the total POPULATION in the country. Distance from the
counterinsurgent’s home country to the conflict country (DISTANCE) has a
small but significant positive effect, which also corresponds to the theoretical
expectations.
Main independent variable

For all cases of insurgency including expeditionary wars, the TBI was calculated for the country in which the actual conflict took place. It is strongly and robustly associated with rebel victory. As explained above, TBI values close to 1 indicate that the state has the entire population in the vicinity of the capital city. Values close to 0 correspond to situations in which the state must reach out to its most remote corners to reach the bulk of the population. Therefore, the large, negative estimate and the small standard error support the theoretical expectation and hypothesis 3.1a. The TBI remains highly significant with the expected negative sign for all fitted models.

Predictive performance

The regression models were used to predict probabilities for each possible outcome of a given war. By taking the maximum of these probabilities as the model's prediction, we can make a comparison with the empirical record. The absolute value of the mismatch between prediction and empirical record provides a simple distance metric that still takes into account the ordinal scale of the dependent variable. This simple deviation can be formalized as $\sum |Y - \hat{Y}| / N$ and corresponding results are also shown in table 7.1. Although it is robustly associated with outcome, the TBI does not improve the predictive capabilities of the models in-sample or out-of-sample. For the out-of-sample tests, leave-one-out cross-validation was performed: The model was fitted on N-1 observations and then used to predict the dependent variable of the missing observation. Again, ordinal deviation was chosen as the measure of success and the process was repeated for all observations. As one can see in table 7.1, the baseline model 1 performs quite well in-sample, with an average deviation score of 0.4462. Only model 5 that includes the TBI is slightly better, while the remaining models are worse. Surprisingly, the minimal model 2 performs best out-of-sample, followed by model 5. In-sample and out-of-sample, the TBI improves statistical predictions. Based on model 5, the effect of the TBI variable on incumbent and insurgent victory is visible in figure 7.2.
values observed in the empirical sample have an effect on the predicted probabilities, but the corresponding error bands are comparatively large for low TBI cases. For high TBI cases, the error bands are much smaller. This implies that uprisings in countries with population concentrations near the capital city are almost twice as likely to be defeated than uprisings in countries with more even population distributions (i.e. TBI values around 0.5). The uncertainty in military outcomes for cases with even population distributions is regrettable, but it is somewhat in line with the simulation results presented in chapter 4: For medium TB settings, the outcomes of the simulations were mostly uncertain, while higher and lower settings reliably led to clear outcomes. Interestingly, the lowest TBI in the conflict sample is India (0.487) and the lowest TBI value globally is associated with Namibia (0.398). With these values well above a theoretically possible value slightly above 0, no maximally periphery-heavy country can be found in the empirical record.

7.4.3 Analyzing casualties

The second macro-implication of the proposed theory touches upon war severity measured in battle-related casualties. The proposed theory and the derived geographic indicator are to some extent capable of explaining why insurgencies in certain countries are more lethal than others. Generally, regression models based on the Poisson and Negative Binomial distributions are suitable to model dependent count variables. Based on the results of a likelihood-ratio test, a Negative Binomial model was preferred over a Poisson-based approach. The Negative Binomial model is also better suited for modeling the heavy-tailed casualty distribution than a Poisson model (see Winkelmann, 2008, 20).

The casualty data provided by the Correlates of War project and CSCW/PRIO do not cover the entire sample of insurgencies. Reliable numbers on casualties from these two sources could only be obtained for 57 out of the 64 cases under investigation. The coding procedure used Correlates of War data (Singer and Small, 1994, Sarkees, 2000) wherever available (Version 4.1, 24 cases) and the best casualty estimates in the CSCW/PRIO data (Lacina and Gleditsch, 2005)
Figure 7.2: Predicted probabilities for the average country based on model 5. The empirical model predicts a substantial increase for the probability of incumbent victory as a function of TBI. For the probability of insurgent victory, however, a substantial decline for higher TBI values is visible.
to supplement the dataset (Version 3.0, 33 additional cases).

Lacina and Gleditsch (2005) discuss some of the conceptual and empirical challenges that arise in the systematic analysis of casualty figures in war. I analyze battle-related casualties to test the presented theoretical argument, thereby excluding those fatalities that were caused by disease, hunger, and other indirect consequences of fighting Lacina and Gleditsch (2005). The margin of error for battle deaths data is comparatively high and the data can be biased, which follows from the limited ability to cross-validate casualty claims and the impossibility of accounting for all casualties in the turmoil of war (see Oliver and Myers, 1999, Davenport and Ball, 1996, Kalyvas, 2004, Siegler et al., 2008). Other recent studies have found that media-based conflict-events datasets code a representative sample of the conflict events obtained from military records (O’Loughlin et al., 2010). With the reliability of casualty data in question, I tried to use it conservatively and excluded cases where insurgencies took place in the context of wider interstate wars, and dropped observations where no best estimate for at least one of the conflict years was available.

Model 6 in table 7.1 is used again as a baseline model for explaining casualties. Model 7 is a minimal model containing only the two functional forms of the TBI. Model 8 shows results for the full model. Model 9 was chosen to minimize the AIC statistic. In all four models, the dependent variable is the number of battle-related fatalities for each war.

Control variables

As visible in the regression table 7.1, several control variables are significantly associated with battle-related casualties. The REGIME variable is negatively associated with casualties (weakly significant), indicating that more democratic regimes are likely to be involved in less severe counterinsurgency wars (see Lacina, 2006). External SUPPORT is robustly associated with higher casualty figures, which corresponds to the intuition that fueling the conflict from the outside leads to more casualties. POWER, reflecting the state’s military capabilities, is not robustly related to casualties, but is generally estimated to
be positive. The *ENERGY* variable has a negative estimate, but statistical significance is not reached. *OCCUPATION* is estimated to be negative and is found to be significant in all models. Possibly, insurgencies against occupying forces leave the incumbent with a natural “exit strategy” in terms of allowing for a retreat from the theater. Domestic incumbents might not have this option, which leads to more devastating wars within the sample.

*ELEVATION* and *DISTANCE* do not seem to have a large effect on casualties. The estimates for *COLDWAR* are positive and highly significant. The Cold War witnessed committed support to both insurgents and incumbents by the superpowers, resulting in a series of intense conflicts such as the struggle for South Vietnam and the Afghan civil war following the Soviet invasion. *MECH* – the degree of mechanization of the armed forces – generally has a positive effect on casualties. As Lyall and Wilson (2009) point out, highly mechanized armies are less capable of forming ties with the civilian population. This means that their ability to tell friend from foe is less developed and the probability of applying inaccurate violence is higher. The combination of low accuracy in the application of force combined with the high lethality of modern arms probably drive this result. The *HELI* variable does not seem to have a significant effect on casualties.

**Main independent variable**

The quadratic TBI specification is highly significant with a large negative estimate. The inverse U-shaped relationship between TBI and the number of fatalities in insurgencies can therefore be found in the empirical record, as seen in figure 7.3. This finding lends support to hypothesis 3.2.

**Predictive performance**

With regard to casualties, the predictive performance of models 6 to 9 was also evaluated. Average deviation between predicted and observed casualties was used as a criterion. Cross-validation was again performed for out-of-sample testing. As shown in table 7.1, the TBI indicator strongly improves casualty
Figure 7.3: Predicted probabilities for the expected casualties as a function of TBI based on model 8. The dotted lines represent 95% confidence intervals.
predictions. In direct comparison, base model 6 performs worse than the TBI-
only model 7. In-sample, the absolute average deviation in predicted casualties
is 855.2186 for model 6, but only 828.004 for model 7. Out-of-sample, model
6 mispredicts the number of casualties on average with 1430.658 and model 7
with only 860.0593. The lowest AIC is achieved with model 9 which includes
the quadratic form of the TBI. In summary, the lowest average prediction
errors both in-sample as well as out-of-sample are achieved with models that
include the TBI.

Residual analysis

Table 9.9 in the supplementary information to this chapter shows the results
of a residual analysis for casualties and outcomes. Analyzing residuals is ben-
eficial, as the results indicate which subset of cases is best captured by the
presented models. The results are sorted by residuals for the prediction of
casualties. The table also shows deviations between actual and predicted out-
comes according to the introduced ordinal scale. The table conveys an intuition
for the abilities and limitations of the econometric analysis and I will briefly
discuss cases that represent both. At first glance, the overall prediction of or-
dinal outcomes seems to work rather well: In 34 of the 57 cases, the predicted
outcomes correspond to the the actual outcomes. In 27 cases, a deviation of
one between actual and predicted outcomes was found. In only one case does
the model predict the opposite of the actual outcome: the Cambodian civil
war. The casualty model also underpredicts the number of casualties at a
margin of almost 61,000 (see row 6) as well as the severity of the Cambodian
insurgency after the Vietnamese invasion (row 2). In these cases, the deviation
between actual and predicted severity can be explained by the extraordinary
historical context of the Vietnam War and the Cambodian genocide separating
the two conflicts temporally. The underlying theory assumes a peripheral in-
surgent movement and the conventional military forces to be the main dyad in
the conflict. The political violence that engulfed much of southeast Asia after
the start of US combat operation in Vietnam in 1965 deviates from this as-

sumption. In both Cambodian cases in the sample, superpowers were involved in aiding both sides. Although I control for external rebel support in the regression analysis, the corresponding binary indicator cannot account for the magnitude of the support in these cases. Two cases that are almost perfectly predicted are the insurgencies in Pakistan (rows 22 and 24). In both cases, and especially with regard to the multi-staged Balochistan conflict, the theoretical argument closely matches the empirical reality: a peripheral uprising in Balochistan fighting the geographically remote government in Islamabad. The model correctly predicts a ceasefire to result from these cases, and the actual and predicted severities match closely. The case that seems most difficult to predict (in terms of casualties) is Afghanistan (rows 1 and 56). This effect might nevertheless be due to the coding choice by Lyall and Wilson (2009): The 1979 civil war before the Soviet invasion and the 1980-1989 anti-Soviet insurgency are coded as separate conflicts. The severity of the first is overestimated by the statistical model, while the severity of the second is underestimated. If those geographically congruent and temporally adjacent wars were combined in the sample, the corresponding prediction would be more accurate.

7.5 Discussion

In this chapter, I have introduced a geographic indicator that expresses the concentration of population within entire countries. For countries that experienced insurgencies between 1970 and 2010, the indicator correlates with the levels of battle-related casualties and the military outcomes of the corresponding wars. The estimated effects correspond to the theoretically derived and numerically simulated expectations. This insight completes the empirical investigation by having shown that a direct link exists between geography, outcome, and casualties in such conflicts.

More specifically, this analysis supports hypothesis 3.1a, stating that distance from the population has a negative effect on insurgent success. Apart from this indicator, institutional and military variables are also robustly re-
lated to outcomes in insurgencies, as Lyall and Wilson (2009) have shown. The results therefore should not be read as a geographic determinism, but a robust probabilistic tendency in line with the theoretical expectations.

With regard to casualties (hypothesis 3.2), the proposed theory has passed another empirical test. The characteristic inverse U-shaped relationship between the territorial balance indicator and casualties can be found in the empirical record, which corresponds to the expectation that civilians are alienated and reactively mobilized by both sides if the bulk of the population resides at medium distances from the capital city. Tests of predictive accuracy show that – within limits – the fates of the most frequent types of civil wars and their severities can be predicted ahead of time, based solely on the introduced measure of population imbalance and distance to the host nation of the counterinsurgency campaign. This insight is important, as the support to challenged governments or active rebel organizations is usually discussed in terms of normative commitment or strategic alliances, but not with regard to the expectable consequences.

Apart from corresponding to the theoretical expectations, the predicted effects also match the observed frequencies in outcomes and casualties in the simulation model with regard to their functional forms. Territory balance in the simulation and the corresponding indicator in the empirical analysis have a negative effect on the probability of insurgent victory and a negative quadratic effect on casualties (compare the empirical effects to the simulated ones in figure 4.3 on page 62). In summary, empirical support for a macro-level association between geographic conditions and aggregate outcomes of uprisings has been found that corresponds to the Lockean simulation results.
Binomial regression models were estimated to test the relevance of the proposed indicator.

Table 7.1: Results from the empirical analysis of outcomes and severity of insurgencies. Five Ordered Logistic and four Negative Binomial regression models were estimated to test the relevance of the proposed indicator.
8 Discussion, conclusion, and outlook

This study set out to provide a unified model of insurgency that helps us to understand, model, and even predict the overall severity and the type of termination of such conflicts. Three research questions had to be answered to clear the way for a coherent theory with predictive capabilities. First, the determinants of indiscriminate violence had to be analyzed, a research question most prominently discussed by Kalyvas (2006). Deviating from Kalyvas’ explanation, I have proposed a simple distance-decay mechanism largely rooted in exogenous geographic conditions rather than endogenously changing levels of military control. The corresponding empirical findings in chapter 5 confirm the theoretical expectation that violence becomes more indiscriminate as the distance to the perpetrators’ power center increases. Therefore, in addition to the initial motivations of rebel organizations (Weinstein, 2007), levels of military control (Kalyvas, 2006), and competition over resources (Metelits, 2010), the geographic constellation of the actors’ power centers needs to be taken into account.

The second question investigated in this study concerns the effects of indiscriminate violence in civil wars. Simplistic deterrence reasoning asserts that higher levels of military force will lead to more compliance among the civilian population (for an overview see Martinez and Morgan, 2011). While such reasoning has been challenged theoretically for a long time (see Ellsberg, 1970, Mason and Krane, 1989, Kilcullen, 2009), attempts to refute such claims empirically face methodological challenges. Most importantly, the absence of natural
spatial units of analysis in conflict events data complicates inferential analyses. Chapter 6 introduced a methodological setup directly tailored to finding causal relationships in conflict events data. In Monte Carlo simulations, the method demonstrated its ability to reliably recover the systematic effects of specific event types on subsequent events. The great advantage of this method over comparable approaches (Kulldorff, 1997, Braithwaite and Johnson, 2012, Linke et al., 2012) is that it avoids the unrealistic assumptions underlying simulated baselines and directly reveals the relative effects of indiscriminate violence compared to selective violence under otherwise similar conditions. I also applied this methodology to study the effects of incumbent and insurgent violence in Afghanistan and found that indiscriminate violence systematically affects civilian collaboration with US forces: Indiscriminate incumbent violence leads to fewer instances of civilian assistance to US forces than selective incumbent violence. Indiscriminate insurgent violence has the opposite effect and leads to more instances of civilian assistance to US forces. This effect is in line with the Lockean assumption that unjustified violence and destruction fuel resistance.

If these two effects are combined, an integrated theory of geography, violence, and mobilization in insurgencies can be derived. The actor whose power center lies closer to the bulk of the population applies violence more selectively than the adversary. This actor receives stronger public support in reaction to the indiscriminate violence used by the adversary. This effect can help the actor succeed even against a militarily superior enemy. Unlike the “greed”-based explanations for participation in civil conflict, this theory stresses the role of past experiences instead of future gains for individual participation in political violence.

The severity of insurgencies also hinges on the spatial constellation of the actors: If one actor is decisively closer to the bulk of the population, the civil conflict will terminate without much bloodshed as the mobilization and support against the sole perpetrator of indiscriminate violence quickly tips the strategic balance. If both actors are remote from the bulk of the population,
repeated cycles of indiscriminate violence and reactive mobilization drive the casualty figures upward. This effect was simulated in an agent-based model in chapter 4. Chapter 7 introduced an empirical indicator that expresses proximity of the military actors to the population in conflict-torn countries. This indicator is significantly associated with both the military outcomes and severities of insurgencies. Probabilities for civil war outcomes and the overall levels of casualties were predicted based on the econometric models. The functional forms of the relationships between this indicator and the dependent variables correspond to the predictions of the agent-based simulation for Lockean parameter settings. This insight strengthens the claim that the presented effects on the micro-level scale to the conflict level and substantially affect outcomes and casualties. A corresponding residual analysis helped to express the performance of the models across different cases. Insurgencies in the context of wider wars and strong external involvement are naturally harder to predict than cases that take place along a single incumbent-insurgent dyad. With these restrictions in mind, the presented model can nevertheless generate informed expectations about the aggregate properties of insurgencies.

From a theoretical standpoint, the presented results provide a possible answer to the introduced puzzles in chapter 2: Insurgencies became more successful during the twentieth century, while increased mechanization has dramatically increased the power projection capabilities of state actors. This seeming paradox is easily explained if the presented effects are taken into account. While states have acquired the ability to apply more violence over longer distances, their ability to apply violence accurately, i.e. conditionally on behavior or affiliation, has increased at a much smaller rate. The use of legitimate violence requires conditionality and proportionality, which is usually implemented through an institutional and legal framework that establishes guilt based on forensic evidence, witness testimony, and specific laws under the presumption of innocence. Advancing weapons technology towards precision strikes does not resolve this problem fundamentally. While guided munitions generally hit their intended target, the selection of the target remains decisive. As long as
the decision to attack is based on behavioral signatures and aerial footage, errors will be made.

In the twentieth century, insurgencies might simply have become more successful as governments started to rely more on heavy arms in fighting off internal challengers. The results of this study suggest that such behavior generates increasing support for the adversary that can prove decisive in population-centric conflict. Demonstrating how such an effect can scale to the macro-level, the introduced simulation model shows a clear negative correlation between the number of suffered casualties and the chances of winning the war as long as reactive mobilization takes place. Only for settings without reactive mobilization can a positive correlation be found.

The second puzzle is how insurgents can mobilize despite the smallest chances of success or even survival for the individual combatant. A materialist reading of the collective action problem suggests that the willingness to take considerable risks depends on the expectations of considerable material rewards. But non-material incentives – such as avenging civilian casualties – can also motivate combatants. This motive is arguably strongest if violence was applied against innocent bystanders, i.e. indiscriminately. Even if the chances of settling the score with the individual perpetrator are extremely remote, this irrational desire for revenge can be an overriding motive (see Pinker, 2011, 530). This mechanism provides a possible explanation for the second theoretical puzzle, i.e. the question of why combatants accept a high probability, or even the certainty of death in the pursuit of armed conflict.

In modeling how this effect scales to the macro-level, one needs to account for changing individual utilities as a function of previous interactions, which is difficult to do in game theory. In the framework of agent-based simulation, however, changes in individual loyalties in response to indiscriminate violence are easily account for (see chapter 4). The simulated macro-level predictions from chapter 4 correspond to the results of the econometric analysis in chapter 7 with regard to their functional forms.

While the results of the empirical studies are encouraging, it is important
to note that they are still limited by data availability. The reactive patterns have so far only been established for Afghanistan in chapter 6. More fine-grained data are necessary to consolidate these insights. Nevertheless, several recent studies have found support for mobilization in reaction to indiscriminate violence across several wars (Downes, 2007, Kalyvas and Kocher, 2009, Condra and Shapiro, 2012, Linke et al., 2012). Moreover, the introduced methodology in chapter 6 will be made available as a package in the programming language R, clearing the path for future studies on reactive dynamics in conflict events data. Therefore, the presented results provide a formidable point of departure for future studies, both theoretically and empirically.

Beyond basic research, the study also has wider implications for policy. In order to discuss these in more detail, I will try to extrapolate the basic geopolitical trends that are likely to shape conflicts in the foreseeable future. While the 9/11 wars are ending, a new wave of conflicts have emerged: the uprisings in the Greater Middle East. Libya and Syria have experienced full-blown civil wars with tens of thousands of casualties. Mali has seen an intensification of its long-standing Tuareg insurgency and the recent surge of violence in Egypt highlights the possibility of military escalation. Future international interventions in such conflicts seem very likely. Against the backdrop of tight national budgets and increasingly capable unmanned aerial vehicles, covert targeted airstrikes and material assistance to strategic allies may be increasingly applied as a viable solution to contain uprisings. While “targeted killings” can be carried out with great precision and technical sophistication, the basic insights generated by this study suggest that the underlying logic of killing combatants to deter followers is fundamentally flawed if violence is applied unconditionally. The great precision and sophistication of such “targeted killings” often obscure their limited accuracy: Up to 50% of the US drone strikes in Pakistan in 2009 and 2010 were so-called “signature strikes”.¹ In signature strikes, in-

individuals are killed based on their behavioral patterns, their interaction with suspected combatants, and their supposed affiliation with insurgent or terrorist organizations. The identity of these individuals, the exact type and magnitude of their criminal or para-military activities is often unknown. The basic principle of (capital) punishment being applied conditionally based on individual misconduct – the very backbone of social contract theories from Locke to Rawls – is violated in these attacks. Advanced sensor technologies, precision-guided munitions, and minimized military losses cannot detract from the reality of such operations. Far from home, with minimal knowledge of the intricate human and physical geography, technocrats kill and command to kill in pursuit of an invisible enemy, thereby fueling the rage and desperation of victims and their kin. More than a decade after the 9/11 attacks, the Authorization for Use of Military Force Against Terrorists (AUMF) remains in place and continues to remove and simultaneously generate enemies of the United States. Military success metrics, constructed around counts of sorties, bodies, and captured combatants and arms, give false indications of the trajectories of irregular conflicts. Simplistic deterrence reasoning makes false predictions for the effects of violence in civil wars. If violence is applied indiscriminately, reactive mobilization for the strategic adversary is likely, which implies that tactical victories and strategic defeat can go hand in hand.

Another implication of this study is that population imbalances crucially determine outcomes and casualties in insurgencies. With some caution, this indicator (or more elaborate measures of population imbalance) could be used to forecast conflict scenarios as insurgencies unfold. If high levels of casualties and a political settlement of the conflict seem probable, peace-keeping interventions should be considered with urgency by the international community. Of course, the presented theory cannot be read deterministically. Instead, cautious forecasts of the trajectories of ongoing civil conflicts and tested assumptions of the processes that drive them seem possible on the basis of the presented results. Moreover, the study lends support to an informed rejection of deterrence-based theories in the analysis of civil conflicts. In light of
advancements in understanding and modeling civil conflict, political decision
making must not be driven by commitment to actors, values, and ideologies
alone. The price in human suffering that results from both interventions and
non-interventions can be estimated to some extent. Ignoring this possibility
means an increased risk of destructive and unnecessary military disasters re-
peating themselves.
9 Supplementary information

9.1 Supplementary information for chapter 4

The robustness of the presented simulation results was assessed for both the
Lockean and the Hobbesian scenarios. In a first set of runs (the upper rows
in figures 9.1 and 9.2 respectively), all parameters were held constant at their
means while territorial balance was systematically varied. Consequently, the
resulting effect for outcome (left column) and casualties (right column) match
those presented in chapter 4. In the bottom rows of figures 9.1 and 9.2, the
effects of varying initial loyalties between 0.4 and 0.6 can be seen. This vari-
ation has a profound influence on the Lockean scenario in terms of watering
down the substantial effect and widening the confidence bounds. Interestingly,
no similar effect can be observed for the Hobbesian scenario. This is easily
explained by the different mechanisms that decide over military success and
failure in the two scenarios. In the Lockean case, the winning side benefits
from high levels of reactive mobilization against the adversary. At the same
time, the winning side suffers higher casualties, because combatants are less
likely to attack while civilian agents are in their vicinity. The resulting paradox
is that the winning side is tactically defeated due to less aggressive behavior
in the field, while it strategically succeeds by compensating for its losses with
newly mobilized combatants. In this situation, the number of casualties for
either side correlates positively with its chances of winning. Initial loyalties
strongly influence this mechanism. If loyalties are against one side, it takes
more instances of indiscriminate violence to reactively mobilize civilians.

Under Hobbesian circumstances, victory comes about differently: As indis-
criminate violence pays off for the perpetrator, attacking either combatants of civilians is a win-win situation. Either enemy combatants are killed, or additional combatants are mobilized for the perpetrator after they witness indiscriminate violence. Under these circumstances, the number of casualties for either side correlates negatively with its chances of winning. Initial loyalties do not play such an important role in this situation. Instead of victory resulting from reactive mobilization, directly inflicted casualties on the adversary bring about victory. The more aggressive side wins under such circumstances, largely independent of initial loyalties (see the bottom row of figure 9.2).
Figure 9.1: Robustness check for agent-based model. In the top row, the Lockean setting (positive alienation factor) was used while all other parameters were held constant at their means. In the bottom row, the initial loyalties were varied from 0.4 to 0.6. It is clearly visible how the effect of territorial balance is “watered down” as variation in initial loyalties is introduced: The magnitude of the effect becomes smaller and the confidence intervals become larger.
Figure 9.2: In a second robustness test, the Hobbesian setting (negative alienation factor) run with and without variation in initial loyalties. Interestingly, the introduced variation does not have the same effect as in the Lockean scenario: the effect remains strong and the confidence bands narrow in the bottom row.
Table 9.1: Regression results from the SIGACT analysis. The estimated models predict indiscriminate violence as a function of distance to the capital and show a positive effect on incumbent indiscriminate violence and a negative effect on insurgent indiscriminate violence. In this case, only direct and indirect fire were used to code the event categories.

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Insurgent indiscriminate</th>
<th>Incumbent indiscriminate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Dist. Pak. (km)</td>
<td>-0.003***</td>
<td>-0.003***</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>Dist. Kabul (km)</td>
<td>-0.003***</td>
<td>-0.003***</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>Elevation</td>
<td>-0.0003***</td>
<td>-0.0003***</td>
</tr>
<tr>
<td></td>
<td>(0.00004)</td>
<td>(0.00004)</td>
</tr>
<tr>
<td>Population</td>
<td>0.00001***</td>
<td>0.00001***</td>
</tr>
<tr>
<td></td>
<td>(0.00000)</td>
<td>(0.00000)</td>
</tr>
<tr>
<td>Line-of-sight</td>
<td>-0.0003**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td></td>
</tr>
<tr>
<td>GECON</td>
<td>-0.476***</td>
<td>-0.476***</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Urban dist.</td>
<td>0.001***</td>
<td>0.001***</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>Landcover</td>
<td>0.197***</td>
<td>0.192***</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Prev. violence</td>
<td>-0.017***</td>
<td>-0.018***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.714***</td>
<td>-0.653***</td>
</tr>
<tr>
<td></td>
<td>(0.144)</td>
<td>(0.140)</td>
</tr>
</tbody>
</table>

Observations: 22,192 22,192 1,193 1,193
Log Likelihood: -12,324.430 -12,326.610 -705.659 -707.187
Akaike Inf. Crit. 24,668.870 24,671.220 1,431.318 1,426.374

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

9.2 Supplementary information for chapter 5
Table 9.2: Robustness test with the casualty counts of the SIGACT data. Note the weak quadratic relationship between distance from capital that corresponds to the findings from the GED analysis.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dist. Kabul (km)</td>
<td>$-0.002^{**}$</td>
<td>$-0.003^{**}$</td>
<td>$-0.005^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Dist. Kabul$^2$</td>
<td>0.00000*</td>
<td>0.00000***</td>
<td>0.00001***</td>
</tr>
<tr>
<td></td>
<td>(0.00000)</td>
<td>(0.00000)</td>
<td>(0.00000)</td>
</tr>
<tr>
<td>Friendly cas.</td>
<td>0.367***</td>
<td>0.369***</td>
<td>0.331***</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.043)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>Enemy cas.</td>
<td>0.025**</td>
<td>0.025**</td>
<td>0.055***</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Population</td>
<td>0.00004***</td>
<td>0.00004***</td>
<td>0.00004***</td>
</tr>
<tr>
<td></td>
<td>(0.00000)</td>
<td>(0.00000)</td>
<td>(0.00000)</td>
</tr>
<tr>
<td>Landcover</td>
<td>0.322***</td>
<td>0.322***</td>
<td>0.322***</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.026)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>GECON</td>
<td>$-0.129^*$</td>
<td>$-0.128^*$</td>
<td>$-0.129^*$</td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
<td>(0.066)</td>
<td>(0.066)</td>
</tr>
<tr>
<td>Pashtun</td>
<td>$-0.057$</td>
<td>$-0.057$</td>
<td>$-0.057$</td>
</tr>
<tr>
<td></td>
<td>(0.093)</td>
<td>(0.093)</td>
<td>(0.093)</td>
</tr>
<tr>
<td>Hazara</td>
<td>$-0.879^{***}$</td>
<td>$-0.879^{***}$</td>
<td>$-0.879^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.331)</td>
<td>(0.331)</td>
<td>(0.331)</td>
</tr>
<tr>
<td>Constant</td>
<td>$-3.964^{***}$</td>
<td>$-4.002^{***}$</td>
<td>$-0.575^{***}$</td>
</tr>
<tr>
<td></td>
<td>(0.315)</td>
<td>(0.300)</td>
<td>(0.122)</td>
</tr>
<tr>
<td>Observations</td>
<td>28,919</td>
<td>28,919</td>
<td>28,920</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>$-11,855.140$</td>
<td>$-11,858.240$</td>
<td>$-11,995.720$</td>
</tr>
<tr>
<td>$\theta$</td>
<td>$0.037^{**}$ (0.001)</td>
<td>$0.037^{**}$ (0.001)</td>
<td>$0.033^{***}$ (0.001)</td>
</tr>
<tr>
<td>Akaike Inf. Crit.</td>
<td>23,730.290</td>
<td>23,732.480</td>
<td>24,001.450</td>
</tr>
</tbody>
</table>

*Note*: $^*$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01
9.3 Supplementary information for chapter 6

Monte Carlo simulations

This section presents an evaluation of the performance of MWA based on simulated events data. Artificial data is used to maximize the transparency of the setup and generate benchmarks under controlled conditions. The constructed data features clear causal patterns, but also random noise that could be expected in any empirical application. Three scenarios will be considered in this section. First, as a proof-of-principle, a predefined pattern is retrieved under ideal conditions: Under otherwise comparable circumstances, treatment and control events that are separated in time and space are analyzed. Second, data with increasingly stronger overlaps of interventions are analyzed to illustrate the resulting biases. Finally, the proposed remedy in terms of matching on SUTVA violations is tested. This simple setup is sufficient to highlight the properties and performance of the method. Variations in exposure and momentum are also included, thereby increasing the complexity and realism of the artificial data.

Data generating process

In order to emulate some of the empirical complexity of events data, artificial samples with three types of events were constructed. One type of event represents the “dependent” category and changes in the occurrence of this event after interventions are the quantity of interest. The other two types are intervention events, labeled “treatment” and “control” events in compliance with the matching terminology.

To emulate an empirical setup, “dependent” events were scattered around treatment and control events. Specifically, they reflect an increase of one event in the dependent category after treatment is applied, while the number of dependent events following controls remains unchanged. They also feature variation both in the total numbers of dependent events and in the trends in the dependent variable before the intervention.
For each episode two stylized confounding factors were assigned which were simply numerical values drawn from the same uniform distributions. For the simulation, the potential effects of confounding factors on the probability of treatment being applied were ignored, since they would be mitigated by matching if they were present.

Artificial intervention events were distributed over a geographic region within 2 by 2 degrees around the equator and the prime meridian, which corresponds to an area of roughly 220 km by 220 km. Figure 9.3 visualizes the generic spatial setup of the test dataset. By varying the simulated time period the density of events in the artificial setup was varied to generate situations in which interventions overlap.

![Figure 9.3: Map of the simulated data distributed over the region within the 1st degree latitude North and South and the 1st degree longitude East and West. This area that corresponds to roughly 220 km by 220 km. This generic spatial setup is used for all Monte Carlo simulations.](image)

The desired causal effect is modeled in two steps. Events of the “dependent” category are placed prior to interventions such that they exhibit different increasing or decreasing trends. Dependent events following interventions are placed at fixed temporal and spatial distances from the trigger. The data used in this study contains 100 such constructed episodes centered around “controls” and 50 episodes centered around a “treatment”. This imbalance is intentionally
chosen to emulate the complications of empirical data.

Simulation results

The MWA method generates estimates for the treatment effect for a number of spatial and temporal window sizes. This output can be graphically communicated based on a contour plot: the lighter the color, the larger the effect. The significance level of the estimated treatment effect is indicated by shading out some of the estimates: No pattern corresponds to $p<0.05$ for the treatment effect in the DD analysis, which means that the effect is substantively interpretable. By convention, estimates are called significant if they have at most a 5% chance of arising under the null hypothesis, which is in this case the assumption that no treatment effect exists. Dotted lines indicate $p$-values between 0.05 and 0.1 and full lines indicate $p>0.1$, i.e. weakly and not significant results respectively. The cells indicating effect size and significance level are arranged in a table where each field corresponds to one specific size of the spatiotemporal cylinders (see figure 9.4).

In a first simple test case, non-overlapping spatio-temporal cylinders were analyzed. Specifically, reactive patterns in which the treatment effect is an increase in one dependent event at a distance of eight days and eight km were constructed. This increase does not occur in the control group. The resulting episodes are randomly distributed over a 20-year period and geographically arranged, as exemplified in figure 9.3. Figure 9.4 nicely summarizes the results: For small spatial and temporal cylinder sizes, for example 2 to 6 days and 2 to 6 km, the estimated treatment effect is 0. Naturally, these areas of the plot are shaded and thereby marked as not significant. This non-result comes about because treatment and control events do no cause different levels of dependent events at such distances from the intervention. Treatment and control only differ at a fixed distance of 8 days and 8 km from the intervention. At this distance, one additional reactive event was placed behind every treatment event. This leads to a significant estimate at 8 days and 8 km in the contour plot. Please note that higher levels of aggregation in the upper right corner
of the plot also show the estimated effect at 1. This is because the simulated effect can also be detected when larger cylinder sizes are chosen.

![Figure 9.4: Estimates and significance levels for an increase of one dependent event within eight days after and eight km from a treatment event. Significance levels are indicated graphically. No pattern corresponds to $p < 0.05$, dotted lines to $p < 0.1$, and full lines to $p > 0.1$.](image)

Summary statistics for the matching procedure for each spatial and temporal window size can be found in figure 9.5. Each cell shows the number of control events ($C$) and treatment events ($T$) both pre- and post-matching as well as the multivariate imbalance measure ($L1$) and the percentage of common support ($\%$). Similar to figure 9.4 the cells are arranged in a table such that their positions correspond to spatial and temporal window sizes. For all window sizes, the percentage of common support increases and the imbalance decreases noticeably with matching; the balance is especially good for the window sizes marked in white where the treatment effect is significant.
Figure 9.5: Matching statistics for the case shown in figure 9.4. The number of control events in the matched sample is indicated with “C”, the number of treatment events with “T”, the multivariate imbalance measure “L1” and the percentage of common support “%”. Window sizes with non-significant estimates are shaded in gray.

### Robustness of the method

A series of tests were run to assess the possible bias arising from overlapping interventions and to demonstrate the effectiveness of the proposed remedies. To enforce overlaps, artificial events data were distributed in the same simulated space as shown in figure 9.6 and with the same reactive pattern as before, but within increasingly shorter time periods (from one year to three weeks). For each time interval, 100 random test datasets were generated and the method was applied to each of them.

Figure 9.6 shows the average estimates and confidence intervals for the causal effect at eight days and eight km as a function of growing overlap in the interventions. The standard matching procedure is compared to a setup where
matching is also performed on SUTVA violations prior to interventions. The “% double treatment” in figure 9.6 indicates the percentage of observations for which at least two treatment events are in the same cylinder. Cases where treatment and control cases overlapped prior to interventions lead to substantively identical results.

![Figure 9.6](image)

Figure 9.6: Average estimates with confidence bounds as a function of the overlap of the spatio-temporal cylinders. The graph shows estimates for MWA, MWA with non-random deletion of overlapping observations, and MWA with matching on counts of previous treatment and control events. Asterisks indicate estimates for all test datasets significant at the 0.05 level and the dotted line marks the true effect.

The figure clearly indicates that all three methods yield correct estimates for small spatio-temporal overlaps (< 10% double treatment). For larger overlaps the estimates of MWA are indeed visibly affected by SUTVA violations. In fact, for overlaps larger than 10%, estimates become less robust (around 50% of the cases for overlaps > 30% are no longer significant). This is similarly true for MWA with deletion of overlapping cases. The most stable results among the three methods are obtained when using MWA with additional matching
on counts of previous treatment and control events. In this case, estimates also remain significant for larger overlaps.

This analysis shows that the method robustly reveals the true causal effect and its spatio-temporal lag for cases with only few overlapping observations (< 10% overlap). In the cases of stronger overlaps, matching on the number of previous treatment and control events improves the accuracy of the estimated treatment effect, which is in line with our theoretical arguments in section 6.3.4, but only to a certain point: Beyond 30% overlap estimates are less robust.

**Placebo tests for SIGACT data**

In order to illustrate the ability of Matched Wake Analysis (MWA) to reveal reactive patterns in conflict events data, two placebo tests were conducted. Figure 9.7 shows the results for two non-relationships in the empirical record. On the left, two criminal and not combat-related event types have been analyzed with regard to their effects on the number of IEDs being subsequently used against US forces. Instances of assassinations or kidnapping are sporadically reported in SIGACT. Since these violent acts cannot be directly attributed to the insurgency, the established “treatment effect” is not significant across the entire range of parameter combinations. Similarly, meetings between US forces and civilians with a focus on improving security versus meetings on economic development do not have a significantly different effect on the subsequent levels of civilian military collaboration. These non-relationships further underline the ability of MWA to tell apart meaningful reactive patterns and noise in SIGACT.
Figure 9.7: Placebo tests: On the left, kidnappings versus assassinations affecting subsequent levels of turn in events are depicted. Meetings between US Forces and civilians with a focus on security versus meetings with a focus on development with regard to their effects on subsequent levels of turned in explosive remnants of war can be seen on the right.

Robustness checks

A series of tests has been performed to ensure that the presented findings are robust. The confounding variables for the matching procedure were changed to demonstrate that the results hold across specifications. All available event categories and the corresponding counts in SIGACT are listed after that. This list serves to further justify the operationalization of selective and indiscriminate violence in the empirical analysis.
Figure 9.8: The analysis from chapter 6 was repeated with alternative confounding factors. In this case, Previous interventions, as well as population counts, the “Pashtun” dummy, spatially disaggregated wealth, as well as distances to Kabul and Pakistan were taken into account. The results remain substantively identical.

Figure 9.9: In this minimal setup, only population figures and previous interventions were used as confounding variables. The results for reactions to indiscriminate insurgent violence remain the same, but the results for incumbent violence become non-significant.
## Codings of the confounding variables

<table>
<thead>
<tr>
<th>Confounding factors for matching</th>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation</td>
<td>Elevation data obtained from the GTOPO30 dataset (Gesch et al., 1999)</td>
</tr>
<tr>
<td>Line-of-Sight</td>
<td>A dataset derived from GTOPO30</td>
</tr>
<tr>
<td>Population</td>
<td>Geo-referenced population counts (CIESIN, 2005)</td>
</tr>
<tr>
<td>GECON</td>
<td>Spatially disaggregated data on wealth (Nordhaus et al., 2006)</td>
</tr>
<tr>
<td>Pashtun settlement area</td>
<td>A dummy indicating whether the event took place in Pashtun areas, derived from Wucherpfennig et al. (2011)</td>
</tr>
<tr>
<td>Hazara settlement area</td>
<td>A dummy indicating whether the event took place in Hazara areas, derived from Wucherpfennig et al. (2011)</td>
</tr>
<tr>
<td>Distance to Pakistan</td>
<td>Distance from the event location to Pakistan, derived from Weidmann et al. (2010)</td>
</tr>
<tr>
<td>Distance to Kabul</td>
<td>Distance from the event location to Kabul, derived from Weidmann et al. (2010)</td>
</tr>
<tr>
<td>Season</td>
<td>An integer indicating the season in which the event took place</td>
</tr>
<tr>
<td>Previous reactive trend</td>
<td>A trend for the reactive event established in the time span under investigation that preceded the event</td>
</tr>
</tbody>
</table>

Table 9.3: Original selection of confounding factors for the presented MWA analysis
List and counts of SIGACT event categories

<table>
<thead>
<tr>
<th>Event Category</th>
<th>#</th>
<th>Event Category</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEETING - DEVELOPMENT</td>
<td>988</td>
<td>PROJECT CLOSEOUT</td>
<td>81</td>
</tr>
<tr>
<td>MEDEVAC (LOCAL NATIONAL)</td>
<td>428</td>
<td>OTHER (HOSTILE ACTION)</td>
<td>417</td>
</tr>
<tr>
<td>RECONNAISSANCE</td>
<td>28</td>
<td>FINANCE</td>
<td>3</td>
</tr>
<tr>
<td>ERW RECOVERED</td>
<td>24</td>
<td>MINE FOUND/CLEARED</td>
<td>396</td>
</tr>
<tr>
<td>INDIRECT FIRE</td>
<td>7229</td>
<td>TRIBAL</td>
<td>7</td>
</tr>
<tr>
<td>AMBUSH</td>
<td>537</td>
<td>PLANNED EVENT</td>
<td>403</td>
</tr>
<tr>
<td>FRAGO</td>
<td>404</td>
<td>COUNTER INSURGENCY</td>
<td>8</td>
</tr>
<tr>
<td>SAFIRE</td>
<td>1695</td>
<td>CRIMINAL ACTIVITY</td>
<td>27</td>
</tr>
<tr>
<td>EQUIPMENT FAILURE</td>
<td>81</td>
<td>REPORTED LOCATION</td>
<td>1</td>
</tr>
<tr>
<td>ARREST</td>
<td>50</td>
<td>TCP</td>
<td>2</td>
</tr>
<tr>
<td>SMUGGLING</td>
<td>22</td>
<td>NEGLIGENT DISCHARGE</td>
<td>19</td>
</tr>
<tr>
<td>SERMON</td>
<td>5</td>
<td>TRIBAL FEUD</td>
<td>12</td>
</tr>
<tr>
<td>SNOW AND ICE REMOVAL</td>
<td>49</td>
<td>COUNTER TERRORISM</td>
<td>1</td>
</tr>
<tr>
<td>GRAFFITI</td>
<td>1</td>
<td>RPG</td>
<td>76</td>
</tr>
<tr>
<td>MEDEVAC PATIENT TRANSFER</td>
<td>160</td>
<td>RECON</td>
<td>33</td>
</tr>
<tr>
<td>DETAINEE TRANSFER</td>
<td>517</td>
<td>UAV</td>
<td>16</td>
</tr>
<tr>
<td>CAS</td>
<td>123</td>
<td>GREEN-GREEN</td>
<td>72</td>
</tr>
<tr>
<td>DEMONSTRATION</td>
<td>236</td>
<td>IDF COUNTER FIRE</td>
<td>5</td>
</tr>
<tr>
<td>DEFECTING</td>
<td>5</td>
<td>GREEN-BLUE</td>
<td>16</td>
</tr>
<tr>
<td>BORDER OPS</td>
<td>11</td>
<td>DRUG OPERATION</td>
<td>6</td>
</tr>
<tr>
<td>CACHE FOUND/CLEARED</td>
<td>2739</td>
<td>POISONING</td>
<td>1</td>
</tr>
<tr>
<td>INTERNAL SECURITY FORCES</td>
<td>2</td>
<td>DELIBERATE ATTACK</td>
<td>69</td>
</tr>
<tr>
<td>CASEVAC</td>
<td>14</td>
<td>BREACHING</td>
<td>2</td>
</tr>
<tr>
<td>AIR ASSAULT</td>
<td>3</td>
<td>COUNTER MORTAR PATROL</td>
<td>7</td>
</tr>
<tr>
<td>SURRENDERING</td>
<td>4</td>
<td>MOVEMENT TO CONTACT</td>
<td>4</td>
</tr>
<tr>
<td>ACCIDENT</td>
<td>834</td>
<td>SECTARIAN VIOLENCE</td>
<td>30</td>
</tr>
<tr>
<td>QA/QC PROJECT</td>
<td>400</td>
<td>DOWNED AIRCRAFT</td>
<td>13</td>
</tr>
<tr>
<td>ELICITATION</td>
<td>1</td>
<td>PSYOP (WRITTEN)</td>
<td>4</td>
</tr>
<tr>
<td>SECURITY BREACH</td>
<td>1</td>
<td>PREMATURE DETONATION</td>
<td>237</td>
</tr>
<tr>
<td>ESCALATION OF FORCE</td>
<td>2267</td>
<td>ATTACK</td>
<td>2267</td>
</tr>
<tr>
<td>SNIPER OPS</td>
<td>154</td>
<td>POLICE INTERNAL</td>
<td>3</td>
</tr>
<tr>
<td>IED HOAX</td>
<td>185</td>
<td>TESTS OF SECURITY</td>
<td>22</td>
</tr>
<tr>
<td>OTHER DEFENSIVE</td>
<td>30</td>
<td>NONE SELECTED</td>
<td>2</td>
</tr>
<tr>
<td>DIRECT FIRE</td>
<td>16286</td>
<td>ANA-ON-ANP</td>
<td>6</td>
</tr>
<tr>
<td>CCA</td>
<td>5</td>
<td>DETAINED</td>
<td>682</td>
</tr>
<tr>
<td>IED AMBUSH</td>
<td>350</td>
<td>VEHICLE INTERDICTION</td>
<td>11</td>
</tr>
<tr>
<td>MEDEVAC</td>
<td>3293</td>
<td>SMALL UNIT ACTIONS</td>
<td>32</td>
</tr>
<tr>
<td>SEARCH AND ATTACK</td>
<td>7</td>
<td>IED FOUND/CLEARED</td>
<td>8369</td>
</tr>
<tr>
<td>INSURGENT VEHICLE</td>
<td>9</td>
<td>UNKNOWN EXPLOSION</td>
<td>155</td>
</tr>
</tbody>
</table>

Table 9.4: Event types and counts in the SIGACT data
<table>
<thead>
<tr>
<th>Event Category</th>
<th>#</th>
<th>Event Category</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAID</td>
<td>44</td>
<td>CHECKPOINT RUN</td>
<td>37</td>
</tr>
<tr>
<td>LOOTING</td>
<td>11</td>
<td>AIR MOVEMENT</td>
<td>8</td>
</tr>
<tr>
<td>KIDNAPPING</td>
<td>109</td>
<td>ENEMY ACTION</td>
<td>13</td>
</tr>
<tr>
<td>SABOTAGE</td>
<td>6</td>
<td>AMNESTY</td>
<td>1</td>
</tr>
<tr>
<td>REFUGEES</td>
<td>12</td>
<td>EVIDENCE TURN-IN/RECEIVED</td>
<td>50</td>
</tr>
<tr>
<td>RESUPPLY</td>
<td>7</td>
<td>BLUE-ON-WHITE</td>
<td>2</td>
</tr>
<tr>
<td>RELEASED</td>
<td>110</td>
<td>NATURAL DISASTER</td>
<td>55</td>
</tr>
<tr>
<td>IED FALSE</td>
<td>550</td>
<td>PROJECT START</td>
<td>88</td>
</tr>
<tr>
<td>OTHER</td>
<td>4684</td>
<td>CONVOY</td>
<td>53</td>
</tr>
<tr>
<td>FOOD DISTRIBUTION</td>
<td>4</td>
<td>BLUE-GREEN</td>
<td>10</td>
</tr>
<tr>
<td>TRANSFER</td>
<td>399</td>
<td>HARD LANDING</td>
<td>9</td>
</tr>
<tr>
<td>IED EXPLOSION</td>
<td>7022</td>
<td>SURVEILLANCE</td>
<td>369</td>
</tr>
<tr>
<td>TURN IN</td>
<td>813</td>
<td>ASSASSINATION</td>
<td>48</td>
</tr>
<tr>
<td>AMF-ON-ANA</td>
<td>2</td>
<td>BLUE-BLUE</td>
<td>18</td>
</tr>
<tr>
<td>ANP TRAINING</td>
<td>282</td>
<td>CORDON/SEARCH</td>
<td>80</td>
</tr>
<tr>
<td>MURDER</td>
<td>99</td>
<td>OTHER OFFENSIVE</td>
<td>132</td>
</tr>
<tr>
<td>PATROL</td>
<td>364</td>
<td>COUNTER NARCOTIC</td>
<td>6</td>
</tr>
<tr>
<td>THEFT</td>
<td>40</td>
<td>EXTORTION</td>
<td>5</td>
</tr>
<tr>
<td>IED SUSPECTED</td>
<td>893</td>
<td>PROPAGANDA</td>
<td>100</td>
</tr>
<tr>
<td>VANDALISM</td>
<td>11</td>
<td>SHOW OF FORCE</td>
<td>2</td>
</tr>
<tr>
<td>RECRUITMENT (WILLING)</td>
<td>1</td>
<td>CLOSE AIR SUPPORT</td>
<td>95</td>
</tr>
<tr>
<td>UNEXPLODED ORDNANCE</td>
<td>2770</td>
<td>ARTY</td>
<td>77</td>
</tr>
<tr>
<td>SUPPORTING AIF</td>
<td>4</td>
<td>ERW/TURN-IN</td>
<td>58</td>
</tr>
<tr>
<td>MEDCAP</td>
<td>160</td>
<td>POLICE ACTIONS</td>
<td>24</td>
</tr>
<tr>
<td>DETAINEE RELEASE</td>
<td>60</td>
<td>MEETING - SECURITY</td>
<td>753</td>
</tr>
<tr>
<td>GREEN-WHITE</td>
<td>6</td>
<td>DETAIN</td>
<td>185</td>
</tr>
<tr>
<td>CARJACKING</td>
<td>30</td>
<td>MINE STRIKE</td>
<td>321</td>
</tr>
<tr>
<td>MEETING</td>
<td>1404</td>
<td>IED THREAT</td>
<td>10</td>
</tr>
<tr>
<td>IDF INTERDICTION</td>
<td>137</td>
<td>NARCOTICS</td>
<td>1</td>
</tr>
<tr>
<td>MUGGING</td>
<td>1</td>
<td>MEDEVAC (OTHER)</td>
<td>64</td>
</tr>
<tr>
<td>INTERDICTION</td>
<td>488</td>
<td>ARSON</td>
<td>41</td>
</tr>
<tr>
<td>PSYOP (TV/RADIO)</td>
<td>2</td>
<td>SUPPORTING CF</td>
<td>15</td>
</tr>
<tr>
<td>REPETITIVE ACTIVITIES</td>
<td>8</td>
<td>BLACK LIST</td>
<td>1</td>
</tr>
<tr>
<td>NBC</td>
<td>1</td>
<td>BLUE-WHITE</td>
<td>6</td>
</tr>
<tr>
<td>COUNTER MORTAR FIRE</td>
<td>41</td>
<td>THREAT</td>
<td>1</td>
</tr>
<tr>
<td>DRUG VEHICLE</td>
<td>2</td>
<td>PSYOP</td>
<td>189</td>
</tr>
<tr>
<td>VOGUE</td>
<td>29</td>
<td>VETCAP</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 9.5: Event types and counts in the SIGACT data (continued)
9.4 Supplementary information for chapter 7

Summary statistics for the main independent variables

Table 9.6 gives an overview of the main independent variables, their distributions and definitions. Please note that all variables except the TBI were taken from a post-1970 subset of the Lyall and Wilson (2009) study.

Ruling out endogeneity in the TBI indicator

Figure 9.10: Densities of TBI values for conflict and non-conflict countries for 1990 (solid lines) and 2010 (dotted lines). The vertical lines mark mean values. Note that conflict countries have a marginally smaller TBI on average, but that there is very little change over time despite the fact that conflicts took place.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
<th>N. Obs</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBI</td>
<td>0.67</td>
<td>0.10</td>
<td>0.49</td>
<td>0.93</td>
<td>65</td>
<td>Main independent variable</td>
</tr>
<tr>
<td>REGIME</td>
<td>-1.54</td>
<td>6.46</td>
<td>-10.00</td>
<td>10.00</td>
<td>65</td>
<td>Polity2 score*</td>
</tr>
<tr>
<td>SUPPORT</td>
<td>1.03</td>
<td>0.85</td>
<td>0.00</td>
<td>2.00</td>
<td>65</td>
<td>Codes external support for the rebels*</td>
</tr>
<tr>
<td>POWER</td>
<td>-1.82</td>
<td>1.63</td>
<td>-4.76</td>
<td>2.82</td>
<td>65</td>
<td>Cumulative index of state capabilities (Log)*</td>
</tr>
<tr>
<td>ENERGY</td>
<td>-0.83</td>
<td>1.69</td>
<td>-6.05</td>
<td>2.33</td>
<td>65</td>
<td>Incumbent energy use divided by population (Log)*</td>
</tr>
<tr>
<td>OCCUPY</td>
<td>0.09</td>
<td>0.29</td>
<td>0.00</td>
<td>1.00</td>
<td>65</td>
<td>Dummy coding if country is was occupied*</td>
</tr>
<tr>
<td>ELEVATION</td>
<td>899.20</td>
<td>1087.31</td>
<td>-20.75</td>
<td>4902.00</td>
<td>65</td>
<td>Average elevation in the conflict area (Log)*</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>471.13</td>
<td>1616.52</td>
<td>1.00</td>
<td>12598.47</td>
<td>65</td>
<td>KM from the incumbent capital to the conflict country*</td>
</tr>
<tr>
<td>COLDWAR</td>
<td>0.60</td>
<td>0.49</td>
<td>0.00</td>
<td>1.00</td>
<td>65</td>
<td>Dummy variable for the Cold War*</td>
</tr>
<tr>
<td>MECH.</td>
<td>2.95</td>
<td>1.07</td>
<td>1.00</td>
<td>4.00</td>
<td>65</td>
<td>Soldiers per motorized vehicle*</td>
</tr>
<tr>
<td>HELI.</td>
<td>0.26</td>
<td>0.44</td>
<td>0.00</td>
<td>1.00</td>
<td>65</td>
<td>Dummy variable for whether incumbent uses helicopters*</td>
</tr>
<tr>
<td>TRADE</td>
<td>-3.46</td>
<td>0.89</td>
<td>-5.99</td>
<td>-0.47</td>
<td>65</td>
<td>Exports+Imports as a share of GDP (Log)*</td>
</tr>
<tr>
<td>LANG.</td>
<td>7.75</td>
<td>7.17</td>
<td>1.00</td>
<td>27.00</td>
<td>65</td>
<td>Number of Languages in the conflict country*</td>
</tr>
</tbody>
</table>

Table 9.6: All variables marked with "*" have been taken from the Lyall and Wilson (2009) dataset. Mech, Polity, Power, and Energy reflect the situation in the last prewar year.
Figure 9.10 shows non-parametric Gaussian kernel estimates for the TBI values based on 1990 and 2010 population figures from the GPW dataset. Please note that in cases with and without conflict between 1990 and 2010, changes in the distributions are marginal. Two possible explanations for this non-effect spring to mind: either internally displaced persons (IDP) do not generally move toward to capital city or the remote periphery, but remain at medium distances from the capital city, or IDPs return to their original settlement areas once the fighting is over. The fact that no systematic variation can be found is important to rule out endogeneity in the empirical analysis. Table 9.7 shows correlations between the main independent variables. Please note that no strong correlation between the explanatory variables can be seen in the data.
Table 9.7: Correlation matrix for the main independent variables. Note that correlations do not exceed 0.62 outside the main diagonal.
Table 9.8: As an additional robustness check, I have calculated binary dependent variable models to predict incumbent success and defeat in insurgencies. Please note that the TBI variable is not significant, but that the sign of the estimate corresponds to the Ordinal Logit model in the thesis. The ordinal information in the dependent variable is therefore necessary for statistically significant results.

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGIME</td>
<td>-0.271***</td>
<td>-0.263***</td>
<td>0.125</td>
<td>0.123</td>
</tr>
<tr>
<td></td>
<td>(0.101)</td>
<td>(0.104)</td>
<td>(0.096)</td>
<td>(0.096)</td>
</tr>
<tr>
<td>SUPPORT</td>
<td>-1.926**</td>
<td>-2.143***</td>
<td>1.135*</td>
<td>1.146*</td>
</tr>
<tr>
<td></td>
<td>(0.754)</td>
<td>(0.822)</td>
<td>(0.613)</td>
<td>(0.642)</td>
</tr>
<tr>
<td>POWER</td>
<td>1.300**</td>
<td>1.665**</td>
<td>-0.993**</td>
<td>-0.843</td>
</tr>
<tr>
<td></td>
<td>(0.545)</td>
<td>(0.715)</td>
<td>(0.489)</td>
<td>(0.579)</td>
</tr>
<tr>
<td>ENERGY</td>
<td>0.440</td>
<td>0.418</td>
<td>-0.557</td>
<td>-0.694*</td>
</tr>
<tr>
<td></td>
<td>(0.298)</td>
<td>(0.302)</td>
<td>(0.340)</td>
<td>(0.388)</td>
</tr>
<tr>
<td>OCCUPY</td>
<td>-17.692</td>
<td>-18.514</td>
<td>4.176*</td>
<td>4.010*</td>
</tr>
<tr>
<td></td>
<td>(2,371.910)</td>
<td>(2,287.469)</td>
<td>(2.144)</td>
<td>(2.254)</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>-0.006*</td>
<td>-0.006*</td>
<td>0.001</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>COLDWAR</td>
<td>0.883</td>
<td>1.100</td>
<td>-1.604*</td>
<td>-1.421</td>
</tr>
<tr>
<td></td>
<td>(0.858)</td>
<td>(0.949)</td>
<td>(0.880)</td>
<td>(0.917)</td>
</tr>
<tr>
<td>HELI.</td>
<td>-5.997***</td>
<td>-6.319***</td>
<td>0.619</td>
<td>1.115</td>
</tr>
<tr>
<td></td>
<td>(2.063)</td>
<td>(2.214)</td>
<td>(1.497)</td>
<td>(1.599)</td>
</tr>
<tr>
<td>LANG.</td>
<td>-0.026</td>
<td>-0.001</td>
<td>0.070</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.072)</td>
<td>(0.057)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>TRADE</td>
<td>-0.491</td>
<td>-0.557</td>
<td>-1.277**</td>
<td>-1.477**</td>
</tr>
<tr>
<td></td>
<td>(0.513)</td>
<td>(0.539)</td>
<td>(0.529)</td>
<td>(0.610)</td>
</tr>
<tr>
<td>log(POPULATION)</td>
<td>-0.494</td>
<td></td>
<td>-0.534</td>
<td>(0.608)</td>
</tr>
<tr>
<td></td>
<td>(0.565)</td>
<td></td>
<td>(0.565)</td>
<td>(0.608)</td>
</tr>
<tr>
<td>TBI</td>
<td>1.677</td>
<td></td>
<td>-5.034</td>
<td>(5.768)</td>
</tr>
<tr>
<td></td>
<td>(4.236)</td>
<td></td>
<td>(5.768)</td>
<td>(5.768)</td>
</tr>
<tr>
<td>Constant</td>
<td>2.895</td>
<td>6.826</td>
<td>-9.959***</td>
<td>-2.610</td>
</tr>
<tr>
<td></td>
<td>(2.627)</td>
<td>(7.014)</td>
<td>(3.545)</td>
<td>(7.972)</td>
</tr>
<tr>
<td>Observations</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-22.775</td>
<td>-22.272</td>
<td>-20.185</td>
<td>-19.615</td>
</tr>
<tr>
<td>Akaike Inf. Crit.</td>
<td>67.550</td>
<td>70.543</td>
<td>62.370</td>
<td>65.230</td>
</tr>
</tbody>
</table>

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

The bivariate relationship between TBI and casualties corresponds very well to the inverse U-shaped functional form discussed in the paper, but the sample is dominated by three cases with more than 40,000 casualties: the civil war in Cambodia, the anti-Soviet insurgency in Afghanistan, and the People’s Mujahideen insurgency in Iran. The inverse U-shaped form is still visible in the data after excluding these cases:

![TBI and casualties for all conflicts](image)

![TBI and casualties for less severe conflicts](image)

Residual analysis

The table below shows residuals for outcomes and casualties for models 5 and 8 respectively. The data was obtained from Lyall and Wilson (2009). Only those cases are shown for which casualty counts could be established.
Table 9.9: Residuals for casualties and outcome based on models 8 and 5 respectively.

Table 9.9 Residuals for casualties and outcome.

<table>
<thead>
<tr>
<th>No.</th>
<th>Incumbent</th>
<th>Insurgent</th>
<th>Start</th>
<th>End</th>
<th>Dev. cas.</th>
<th>Dev. out.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Soviet Union</td>
<td>Afghanistan</td>
<td>1980</td>
<td>1989</td>
<td>-283481</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Cambodia</td>
<td>Khmer Rouge</td>
<td>1978</td>
<td>1992</td>
<td>-218836</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Iran</td>
<td>MEK</td>
<td>1979</td>
<td>2001</td>
<td>-166009</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Lebanon</td>
<td>Various</td>
<td>1975</td>
<td>1990</td>
<td>-83286</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Mozambique</td>
<td>RENAMO</td>
<td>1976</td>
<td>1992</td>
<td>-66069</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>Congo</td>
<td>Cobras, Ninjas</td>
<td>1997</td>
<td>1999</td>
<td>-56896</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>El Salvador</td>
<td>FMLN</td>
<td>1979</td>
<td>1992</td>
<td>-30296</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>Uganda</td>
<td>NRA</td>
<td>1981</td>
<td>1987</td>
<td>-22872</td>
<td>-1</td>
</tr>
<tr>
<td>10</td>
<td>Nicaragua</td>
<td>FSLN</td>
<td>1978</td>
<td>1979</td>
<td>-15338</td>
<td>-1</td>
</tr>
<tr>
<td>11</td>
<td>Tajikistan</td>
<td>UTO</td>
<td>1992</td>
<td>1997</td>
<td>-14847</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>Peru</td>
<td>Sendero Luminoso</td>
<td>1980</td>
<td>1999</td>
<td>-8892</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>Liberia</td>
<td>NPFL</td>
<td>1989</td>
<td>1997</td>
<td>-7462</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>Turkey</td>
<td>Kurds</td>
<td>1983</td>
<td>1999</td>
<td>-6438</td>
<td>-1</td>
</tr>
<tr>
<td>15</td>
<td>Sri Lanka</td>
<td>LTTE</td>
<td>1987</td>
<td>1989</td>
<td>-4536</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>Russia</td>
<td>Chechens</td>
<td>1994</td>
<td>1996</td>
<td>-3285</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>Burundi</td>
<td>FDD</td>
<td>1993</td>
<td>2005</td>
<td>-2311</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>Sierra Leone</td>
<td>RUF</td>
<td>1991</td>
<td>1999</td>
<td>-1463</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>Azerbaijan</td>
<td>Armenia</td>
<td>1992</td>
<td>1994</td>
<td>-848</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>Sri Lanka</td>
<td>LTTE</td>
<td>1983</td>
<td>1987</td>
<td>-580</td>
<td>-1</td>
</tr>
<tr>
<td>21</td>
<td>Liberia</td>
<td>LURD/MODEL</td>
<td>1999</td>
<td>2003</td>
<td>-491</td>
<td>-1</td>
</tr>
<tr>
<td>22</td>
<td>Pakistan</td>
<td>Baluchi</td>
<td>1973</td>
<td>1977</td>
<td>-180</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>India</td>
<td>Sikhs</td>
<td>1984</td>
<td>1994</td>
<td>-40</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>Pakistan</td>
<td>Mohajirs</td>
<td>1993</td>
<td>1999</td>
<td>-17</td>
<td>0</td>
</tr>
</tbody>
</table>

Continued on next page
## Table 9.9 – continued.

<table>
<thead>
<tr>
<th>No.</th>
<th>Incumbent</th>
<th>Insurgent</th>
<th>Start</th>
<th>End</th>
<th>Dev. cas.</th>
<th>Dev. out.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>Papua New</td>
<td>BRA</td>
<td>1988</td>
<td>1998</td>
<td>498</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>Moldova</td>
<td>Dniester</td>
<td>1992</td>
<td>1992</td>
<td>1406</td>
<td>-1</td>
</tr>
<tr>
<td>27</td>
<td>Ivory Coast</td>
<td>PMIC</td>
<td>2002</td>
<td>2005</td>
<td>2087</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>Israel</td>
<td>Palestinian</td>
<td>1987</td>
<td>1993</td>
<td>2588</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>Guinea Bissau</td>
<td>Mil factions</td>
<td>1998</td>
<td>1999</td>
<td>2747</td>
<td>-1</td>
</tr>
<tr>
<td>30</td>
<td>Burundi</td>
<td>Hutu rebels</td>
<td>1972</td>
<td>1972</td>
<td>3027</td>
<td>1</td>
</tr>
<tr>
<td>31</td>
<td>Syria</td>
<td>Muslim Brotherhood</td>
<td>1980</td>
<td>1982</td>
<td>3173</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>Ethiopia</td>
<td>Eritrea</td>
<td>1974</td>
<td>1991</td>
<td>4540</td>
<td>0</td>
</tr>
<tr>
<td>33</td>
<td>Iraq</td>
<td>Kurds</td>
<td>1991</td>
<td>1991</td>
<td>5431</td>
<td>0</td>
</tr>
<tr>
<td>34</td>
<td>Rwanda</td>
<td>ALiR</td>
<td>1994</td>
<td>2000</td>
<td>6755</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>Sri Lanka</td>
<td>JVP</td>
<td>1971</td>
<td>1971</td>
<td>7817</td>
<td>1</td>
</tr>
<tr>
<td>36</td>
<td>Nicaragua</td>
<td>Contras</td>
<td>1981</td>
<td>1988</td>
<td>7868</td>
<td>0</td>
</tr>
<tr>
<td>37</td>
<td>Iraq</td>
<td>Kurds</td>
<td>1980</td>
<td>1988</td>
<td>7958</td>
<td>0</td>
</tr>
<tr>
<td>38</td>
<td>Argentina</td>
<td>ERP</td>
<td>1973</td>
<td>1977</td>
<td>8551</td>
<td>1</td>
</tr>
<tr>
<td>39</td>
<td>Georgia</td>
<td>Abkhazia</td>
<td>1992</td>
<td>1994</td>
<td>8827</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>CAR</td>
<td>Factions</td>
<td>1994</td>
<td>1997</td>
<td>12161</td>
<td>0</td>
</tr>
<tr>
<td>41</td>
<td>Chad</td>
<td>Rebels</td>
<td>1994</td>
<td>1998</td>
<td>13402</td>
<td>1</td>
</tr>
<tr>
<td>42</td>
<td>Bangladesh</td>
<td>Shanti Bahini</td>
<td>1976</td>
<td>1997</td>
<td>14213</td>
<td>-1</td>
</tr>
<tr>
<td>43</td>
<td>Mali</td>
<td>Tuaregs</td>
<td>1989</td>
<td>1995</td>
<td>18374</td>
<td>0</td>
</tr>
<tr>
<td>44</td>
<td>Yemen</td>
<td>Socialist Party</td>
<td>1986</td>
<td>1987</td>
<td>18719</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>Israel</td>
<td>Syria</td>
<td>1982</td>
<td>1982</td>
<td>23237</td>
<td>1</td>
</tr>
<tr>
<td>46</td>
<td>Jordan</td>
<td>Fedeyeen</td>
<td>1970</td>
<td>1970</td>
<td>24853</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>DRC</td>
<td>FLNC</td>
<td>1977</td>
<td>1978</td>
<td>28491</td>
<td>1</td>
</tr>
<tr>
<td>48</td>
<td>Djibouti</td>
<td>FRUD</td>
<td>1991</td>
<td>1994</td>
<td>32299</td>
<td>0</td>
</tr>
<tr>
<td>49</td>
<td>Zimbabwe</td>
<td>ZANU</td>
<td>1972</td>
<td>1979</td>
<td>32770</td>
<td>0</td>
</tr>
<tr>
<td>50</td>
<td>Iraq</td>
<td>Kurds</td>
<td>1974</td>
<td>1975</td>
<td>37958</td>
<td>0</td>
</tr>
<tr>
<td>51</td>
<td>Morocco</td>
<td>Polisario</td>
<td>1975</td>
<td>1989</td>
<td>44045</td>
<td>0</td>
</tr>
<tr>
<td>52</td>
<td>Afghanistan</td>
<td>Taliban</td>
<td>1992</td>
<td>1996</td>
<td>76765</td>
<td>0</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>No.</th>
<th>Incumbent</th>
<th>Insurgent</th>
<th>Start</th>
<th>End</th>
<th>Dev. cas.</th>
<th>Dev. out.</th>
</tr>
</thead>
<tbody>
<tr>
<td>53</td>
<td>Somalia</td>
<td>SSDF</td>
<td>1981</td>
<td>1991</td>
<td>85715</td>
<td>1</td>
</tr>
<tr>
<td>54</td>
<td>Iran</td>
<td>KDPI</td>
<td>1979</td>
<td>1996</td>
<td>93239</td>
<td>1</td>
</tr>
<tr>
<td>55</td>
<td>Chad</td>
<td>Libya/Frolinat</td>
<td>1975</td>
<td>1988</td>
<td>113166</td>
<td>0</td>
</tr>
<tr>
<td>56</td>
<td>Afghanistan</td>
<td>Afghans</td>
<td>1978</td>
<td>1979</td>
<td>224999</td>
<td>-1</td>
</tr>
<tr>
<td>57</td>
<td>Sudan</td>
<td>SPLM</td>
<td>1983</td>
<td>2004</td>
<td>243855</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 9.10: This table shows regression results for outcome and casualties based on a different operationalization of the TBI. Instead of distances to the capital city, distances to the nearest major city were calculated based on Nelson (2008). Please note that the results are substantively identical to the results reported in the thesis based on distances to the capital city. Due to computational limitations, I had to omit the insurgency cases for Russia/Soviet Union, which resulted in 63 instead of 65 observations.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Outcome ordered logistic</th>
<th>Casualties negative binomial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>REGIME</td>
<td>0.128**</td>
<td>0.180*</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>SUPPORT</td>
<td>0.971**</td>
<td>0.944**</td>
</tr>
<tr>
<td></td>
<td>(0.590)</td>
<td>(0.388)</td>
</tr>
<tr>
<td>POWER</td>
<td>-0.503**</td>
<td>-0.347</td>
</tr>
<tr>
<td></td>
<td>(0.231)</td>
<td>(0.222)</td>
</tr>
<tr>
<td>ENERGY</td>
<td>0.220</td>
<td>-0.250</td>
</tr>
<tr>
<td></td>
<td>(0.219)</td>
<td>(0.222)</td>
</tr>
<tr>
<td>OCCUPY</td>
<td>1.964</td>
<td>2.387*</td>
</tr>
<tr>
<td></td>
<td>(1.949)</td>
<td>(1.334)</td>
</tr>
<tr>
<td>ELEVATION</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.0005)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>0.001**</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>COLDWAR</td>
<td>-1.036*</td>
<td>-1.073*</td>
</tr>
<tr>
<td></td>
<td>(0.593)</td>
<td>(0.591)</td>
</tr>
<tr>
<td></td>
<td>-0.022</td>
<td>-0.066</td>
</tr>
<tr>
<td></td>
<td>(0.286)</td>
<td>(0.290)</td>
</tr>
<tr>
<td></td>
<td>1.818**</td>
<td>1.801*</td>
</tr>
<tr>
<td></td>
<td>(0.869)</td>
<td>(0.877)</td>
</tr>
<tr>
<td></td>
<td>0.078*</td>
<td>0.086*</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.043)</td>
</tr>
<tr>
<td></td>
<td>-0.579</td>
<td>-0.614</td>
</tr>
<tr>
<td></td>
<td>(0.419)</td>
<td>(0.413)</td>
</tr>
<tr>
<td></td>
<td>0.164</td>
<td>-0.286</td>
</tr>
<tr>
<td></td>
<td>(0.167)</td>
<td>(0.168)</td>
</tr>
<tr>
<td></td>
<td>-2.903***</td>
<td>-8.103**</td>
</tr>
<tr>
<td></td>
<td>(0.191)</td>
<td>(0.099)</td>
</tr>
<tr>
<td>URBAN TBI</td>
<td>-2.826***</td>
<td>-2.439***</td>
</tr>
<tr>
<td></td>
<td>(0.191)</td>
<td>(0.126)</td>
</tr>
<tr>
<td>URBAN TBI^2</td>
<td>-1.007</td>
<td>-2.763**</td>
</tr>
<tr>
<td></td>
<td>(0.286)</td>
<td>(0.5525)</td>
</tr>
<tr>
<td>Constant</td>
<td>-30</td>
<td>0.517*** (0.082)</td>
</tr>
<tr>
<td></td>
<td>(0.63)</td>
<td>(1.253)</td>
</tr>
</tbody>
</table>

Note: p < 0.05; **p < 0.01; ***p < 0.001.
Bibliography


