How Do You Strike Me?
Decomposing the Determinants of Selective and Indiscriminate Violence in Civil Conflicts*

Gaku Ito†

Abstract

Warring parties choose particular forms of violence in a deliberate attempt to achieve their political objectives in civil conflicts. But why do warring parties employ violence selectively in some locations but indiscriminately in others? Two primary classes of predictors of civil war violence can be found in the literature: the first class includes dynamic factors such as the balance of territorial control endogenous to the conflict process, whereas the second includes relatively static factors such as physical geography that are largely exogenous to the conflict process. This chapter argues that the relative importance of endogenous and exogenous determinants depends on the types of violence applied. Exogenous factors play an important role in predicting indiscriminate violence, because (1) this type of violence is primarily motivated by damage-maximizing incentives, and (2) the locations where warring parties can maximize pain on their opponents are mainly determined by exogenous factors. By contrast, endogenous factors matter in determining the locations of selective violence, because (1) the availability of information needed to apply violence selectively is largely a function of levels of territorial control exercised by the warring parties (Kalyvas, 2006), and (2) the use of violence itself contributes to changes in territorial control. This chapter tests this theoretical argument against the empirical records of insurgent violence in the war in Afghanistan, utilizing a parsimonious agent-based computational model incorporated with precisely geo-referenced data. The simulation results provide empirical support for our theoretical expectations.

---

*Paper prepared for the presentation at the 60th Annual Meeting of the Japan Association of International Relations, Makuhari, Chiba, October 14–16, 2016. Note to the reader: What follows is a summarized version of Chapter 4 of my dissertation. Web appendix is available at {URL}. The full version of the dissertation explores the micro-level determinants of civil war violence and examines the impacts such micro-dynamics have on eventual duration and outcome of civil conflicts. Funding: Financial support from the Japan Society for the Promotion of Science, JSPS (Grant-in-Aid for Scientific Research, KAKENHI, Grant Numbers 16H03589 and 16H06819) is gratefully acknowledged.

†Research Fellow, Center for Transdisciplinary Innovation, National Institutes for the Humanities and Center for Far Eastern Studies, University of Toyama, 3190 Gofuku, Toyama 930–8555. Phone: +81–(0)76–445–6442, Email: gito@eco.u-toyama.ac.jp.

Warring parties deliberately choose particular forms of violence with the aim of achieving their political objectives in civil conflicts (Bueno de Mesquita, 2013; Kalyvas, 2006). But why do warring parties employ violence selectively in some locations but indiscriminately in others? Significant variations are observed in the severity and types of violence within and across civil conflicts. Local distribution of the number and types of perpetrated violence is far from uniform even within a single conflict, as some localities experience severe civilian abuse while other localities are rarely exposed to such victimization during direct military confrontations between armed troops. What drives the spatial variation in the types of violence applied in civil conflicts? Why do the scale and forms of violence vary in civil conflicts?

Utilizing newly available micro-level, precisely-geocoded datasets of civil war battles, scholars have increasingly explored the determinants of insurgent violence in civil conflicts. Based on the difference in target selection, the existing literature regularly employs the conceptual distinction between selective and indiscriminate violence (Kalyvas, 2006, Chap. 6; see also, Ellsberg, 1970; Hechter, 1987; Leites and Wolf, 1970). Selective violence, or the punishment according to individual criteria, refers to violence applied conditional on the past behavior of the targets and is typically observed as violence targeted at collaborators of the opponent. In indiscriminate violence or "reprisals," on the other hand, personalized targeting in selective violence is replaced by collective criteria, typically based on ethnic group affiliation and settled localities, and such instances of violence are often observed as intentional civilian abuse by warring parties during civil conflicts.[1]

As reviewed in detail below, the civil war literature in recent years has increasingly explored the impacts that levels of territorial control (e.g., Kalyvas, 2006), bat-

[1]The types and forms of violence mainly refer to the selectivity of targets independent of the scale of targeting (Kalyvas, 2006). The distinction between selective and indiscriminate violence is analogous to the distinction between selective incentives and collective goods. Selective violence and incentives are provided conditional on the past behavior of individuals, while indiscriminate violence, or collective "bads," and collective goods are distributed on the basis of membership in a group (Hechter, 1987; Kalyvas and Kocher, 2007; Olson, 1965). Rebel groups can employ various tactics to gain civilian support and extract local resources, from the provision of economic incentives and local public goods to coercion and predation (Azam, 2006; Herbst, 2000; Lichbach, 1995). Because indiscriminate violence is often targeted at members of, for example, a specific ethnic group rather than applied completely at random, Steele (2009) proposes the concept of "collective violence" to describe this type of violence. Souleimanov and Siroky (2016) distinguish between "random" and "redistributive" violence, or sub-types of indiscriminate violence. Empirical records of civil war violence in the Chechen wars demonstrate that the instances of these two types of violence have differing impacts on subsequent violent activities of the opponent.

tlefield dynamics (e.g., Hultman, 2007; Wood, 2014a), competition among rebel groups (e.g., Wood and Kathman, 2015), organizational configurations of warring actors (e.g., Humphreys and Weinstein, 2006), relative reliance on local and external sources of support (e.g., Salehyan, Siroky, and Wood, 2014; Zhukov, 2017), and ethnic and physical geography (e.g., Fjelde and Hultman, 2014; Schutte, 2017) each have on the types, frequency, locations, and severity of violence in civil conflicts. Previous empirical studies have demonstrated that these factors, either exogenous or endogenous to conflict dynamics, substantially shape how violence unfolds in the context of civil conflict.

What remains relatively under-investigated in the literature is the relative importance of each class of factors. Perhaps a noteworthy aspect of existing studies is their division of labor, reflecting the prediction targets. Those studies that explore the determinants of selective violence tend to stress the role of factors that are largely endogenous to the conflict dynamics (e.g., territorial control), while those focusing on the determinants of indiscriminate violence typically highlight the role of largely preexisting factors that are often exogenous to the conflict processes (e.g., physical geography). Although these studies offer valuable insights into the possible determinants and mechanisms of civil war violence, we know relatively little about how and why the determinants of selective and indiscriminate violence may differ from each other.

This chapter argues that the relative importance of endogenous and exogenous determinants of violence depends on the types of violence applied. Exogenous factors play an important role in predicting indiscriminate violence, because (1) this type of violence is primarily motivated by damage-maximizing incentives, and (2) the locations where warring parties can maximize their opponents' pain are largely determined by exogenous factors. By contrast, endogenous factors matter in determining the locations of selective violence, because (1) the availability of the information required to apply violence selectively is largely a function of levels of territorial control (Kalyvas, 2006), and (2) the use of violence itself contributes to changes in levels of territorial control.

In order to disentangle the determinants of selective and indiscriminate violence, this chapter employs the empirically-grounded agent-based model developed in the previous chapter. This computational approach enables us to clearly specify the hypothesized mechanisms and generate hypothetical spatial distributions of violence that are directly comparable with the observed records. Because the hypothetical distributions are computationally derived from the computational model, this approach serves as a valuable test of whether theoretical propositions about insurgents' behavior can generate and explain the empirical reality.

The simulation exercise yields two major findings and provides strong empirical support for our theoretical expectations: indiscriminate violence can be predicted well solely by exogenous factors, while endogenous factors, or the recent history of violence that captures the battlefield dynamics and changing balance of territorial control, are vital in predicting where selective violence is applied.

The remainder of this chapter is organized as follows. In Section 1, we examine the recent expansion of the literature on civil war violence, followed by theoretical propositions. The case and empirical data are explained in Section 3, and we propose a parsimonious but empirically-grounded computational model in Section 4. We highlight the empirical results in Sections 5 and 6, and briefly report the results of sensitivity tests in Section 7. We then conclude by offering the scholarly and policy implications of our findings.

1 State of the Debate

The last decade has witnessed a tremendous growth in scholarly understanding of the determinants of violence in civil conflicts. The existing literature demonstrates that levels of territorial control (e.g., Kalyvas, 2006), battlefield dynamics (e.g., Hultman, 2007, 2012; Lyall, 2009; Souleimanov and Siroky, 2016; Wood, 2014a), competition among rebel groups (e.g., Metelits, 2010; Raleigh, 2012; Wood and Kathman, 2015), organizational configurations of warring actors (e.g., Azam, 2006; De la Calle, 2017; Eck, 2014; Humphreys and Weinstein, 2006; Johnston, 2008; Stanton, 2013; Weinstein, 2005, 2007), relative reliance on local and external sources of support (e.g., Ottmann, 2017; Salehyan, Siroky, and Wood, 2014; Toft and Zhukov, 2015; Wood, 2014b; Zhukov, 2017), and ethnic and physical geography (e.g., Balcells, 2011; Di Salvatore, 2016; Fjelde and Hultman, 2014; Schutte, 2017) each invariably influence how violence unfolds in civil conflicts. This section briefly reviews these recent advances in the literature and elaborates the state of the scholarly debate.

1.1 Candidate Determinants of Civil War Violence

Territorial control and conflict dynamics Kalyvas (1999, 2006) brought back into the literature the conceptual distinction between selective and indiscriminate violence proposed by Leites and Wolf (1970). Selective violence involves individual targeting, whereas violence is indiscriminate when targeting is based on collective criteria (Kalyvas and Kocher, 2007, 187–188). The primary predictor of civil war violence in Kalyvas (2006) is the distribution of territorial control. Warring parties employ selective violence in zones of dominant but incomplete territorial control to foster civilian collaboration while deterring support for their opponents. In contrast, the frequency of indiscriminate violence is expected to be inversely related to the level of territorial control. This type of violence, due to the lack of intelligence to discriminate between collaborators of the opponents and innocent civilians, tends to be perpetrated where armed groups have very limited levels of territorial control.

Yet, indiscriminate violence is counterproductive in altering civilian behavior, because the "'innocent' can do little or nothing to escape punishment and the 'guilty' are no more (and sometimes less) threatened" (Kalyvas, 2006, 171). "In a regime of indiscriminate terror," as Kalyvas (1999) argues, "compliance [with the perpetrator] guarantees no security" (251). Due to its counterproductive nature, therefore, indiscriminate violence is expected to be the "product of a lag" and to decline as conflict persists. As warring actors learn the counterproductive nature of the indiscriminate use of violence, they eventually switch to selective violence (Kalyvas, 2006, 172). Utilizing the recently declassified, precisely geocoded records of combat activities from the Hamlet Evaluation System (HES), Kalyvas and Kocher (2009) have examined these theoretical predictions against the observed associations between territorial control and violence in the Vietnam War (see also, Dell and Querubin, 2018; Kalyvas and Kocher, 2007; Kocher, Pepinsky, and Kalyvas, 2011). Consistent with the theoretical claims in Kalyvas (2006), the empirical records show that the locations of selective and indiscriminate violence tend to be separated in space, and highlight the role of territorial control in determining the locations and types of violence perpetrated by warring actors.[2]

Rising battlefield losses and attrition would incentivize warring parties to employ violence indiscriminately, thereby shaping the frequency and manner of violence applied

[2] A related issue in the literature is the effectiveness of selective and indiscriminate violence in mobilizing civilian support and containing the opponents' activities. Kalyvas's theoretical prediction can be summarized as "[t]o be efficient, terror needs to be selective; indiscriminate terror tends to be counterproductive" (Kalyvas, 1999, 251). The empirical results in Dell and Querubin (2018), Kalyvas and Kocher (2007), and Kocher, Pepinsky, and Kalyvas (2011) provide support for the theoretical claim, whereas Downes (2007), Lyall (2009), and Merom (2003) highlight the violence-reducing effect of indiscriminate counterinsurgency campaigns. Toft and Zhukov (2015) stress the conditioning effect of rebels' relative reliance on local and external sources of support.

during civil conflicts (Downes, 2007; Hultman, 2007, 2012; Lyall, 2009; Souleimanov and Siroky, 2016; Wood, 2014a). Building upon the bargaining model of war, Hultman (2007) proposes that recent losses in the battlefield incentivize rebels to target civilians in order to impose political and military costs on the incumbent. The instrumental use of violence against civilians demonstrates rebels' "power to hurt" (Schelling, 1966) and thereby improves their bargaining position against the incumbent (see also, Acosta, 2016; Hultman, 2009, 2012; Stanton, 2013). Wood (2014a) further highlights the conditioning effects of largely static characteristics of rebel groups, such as effective territorial control and sources of rebel financing, on the relationship between rebels' battlefield losses and incentives for civilian victimization.

A related determinant of civil war violence is rebels' inter-group competition over local resources and bargaining power relative to the incumbent (Metelits, 2010; Raleigh, 2012; Wood and Kathman, 2015). Wood and Kathman (2015) contrast dynamic changes in the severity of competition among rebel groups during conflicts with the mere existence of multiple groups. Existing rebel groups are more likely to intentionally target civilians upon the entrance of new groups into the conflict because existing groups may perceive the arrival of new groups as a threat to their control of resources and the expected payoff of winning the conflict. Targeting civilians selectively offers a means to foster civilian collaboration and deter defection, thereby securing their material capability and bargaining power against the incumbent.

Group characteristics and geographic conditions Perhaps a common aspect of these arguments is their focus on dynamic factors that are essentially endogenous to the conflict process, such as the changing balance of territorial control (Schutte, 2017, 381–382). Nevertheless, several empirical studies have examined the role of relatively static factors that are largely, if not completely, predetermined and exogenous to conflict dynamics in altering the frequency and type of violence in civil conflicts.

The internal characteristics of rebel groups are one of these static determinants of civil war violence (Azam, 2006; Beardsley and McQuinn, 2009; De la Calle, 2017; Eck, 2014; Humphreys and Weinstein, 2006; Johnston, 2008; Stanton, 2013; Toft and Zhukov, 2015; Weinstein, 2005, 2007; Wood, 2010, 2014b). For example, Humphreys and Weinstein (2006) posit that high levels of civilian abuse tend to be conducted by warring actors that lack the capabilities to coordinate and police the actions of their members. Armed groups that are ethnically fragmented, tend to rely on material incentives or economic endowments to mobilize participants, and lack credible internal mechanisms for punishing indiscipline, tend to suffer from an inability to monitor their members' actions. Armed groups with such characteristics are therefore expected to be more likely to abuse civilians. The micro-level empirical records of civilian abuse conducted by multiple rebel groups in Sierra Leone confirm these theoretical expectations (see also, Weinstein, 2005, 2007).[3]

Another camp of the literature stresses the role of human and physical geography. Fjelde and Hultman (2014) highlight the role of local ethnic configuration and argue that warring actors often use ethnic affiliation to identify groups with suspected loyalty to the opponents when individual wartime affiliations remain private information. Warring actors, who often depend on civilian support to sustain combat activities, target the suspected enemy collaborators using local ethnic configurations as cues to guide their target selection in order weaken the enemy's capacity. The empirical patterns of civilian abuse in civil conflicts in Sub-Saharan Africa between 1989 and 2009 are consistent with their theoretical claims. In a similar vein, Balcells (2011) argues that indirect violence (violence perpetrated with heavy weaponry) tends to be applied to localities associated with levels of prewar support for the opponent, while direct violence (violence perpetrated with light weaponry) tends to increase with the level of political parity between factions in a locality. The empirical records of violence in the Spanish Civil War (1936–1939) provide support for the posited relationships. Utilizing a novel estimating methodology and survey data in Afghanistan, Hirose, Imai, and Lyall (2017) convincingly demonstrate that village-level pro-government attitudes are followed by an increased risk of insurgent attacks.

While local support for warring parties and, to a lesser extent, local ethnic configurations, can increase over time, geographical conditions such as elevation and distance from national capitals rarely change during the course of conflict. Schutte (2015, 2017) focuses on the role of physical geography, which is almost purely, if not completely, exogenous to conflict processes. Schutte (2017) extends Boulding's (1962) notion of

[3] This logic can be applied to explain the impact of external sources of support on insurgent behavior. Heavy reliance on external, rather than local, sources of support reduces warring actors' need to win the "hearts and minds" of the local civilian population in order to sustain their campaigns and increases the risk of civilian abuse (Salehyan, Siroky, and Wood, 2014; Zhukov, 2017). In a similar vein, Stanton (2013) demonstrates how the size of rebels' civilian constituency influences types of rebel violence. Ottmann (2017) highlight the role of constituency overlap between rebels and the incumbent as well as the monadic civilian constituency in shaping the severity of violence against civilians.

the "loss of strength gradient" (LSG) to explain the quality of targeting and proposes the notion of "loss of accuracy gradient" (LAG). The stylized model predicts that the selectiveness of applied projected violence decays as a function of distance from the warring actors' power centers (e.g., national capitals and rebel bases in periphery) due to the growing inability of the actors to distinguish between collaborators of the opponents and "innocent" locals, or due to the warring parties' "information problem" (Kalyvas, 2006). Empirical analyses using the geocoded event datasets of the ongoing war in Afghanistan and 10 cases of African insurgencies provide support for the notion of LAG.

1.2 Endogenous and Exogenous Determinants of Civil War Violence

The candidate determinants of violence in civil conflicts illustrated above can be thought of as representing a continuum with purely exogenous or static factors at one extreme and purely endogenous or dynamic factors at the other, as depicted in Figure 1. Geographic conditions such as distance to national capitals and elevation are most exogenous to conflict and lie at the left end, while territorial control and battlefield dynamics are largely determined by conflict processes and thus lie at the opposite end of the continuum. Other classes of determinants of violence are located between both ends, as the degree to which these factors are endogenous or exogenous to conflict dynamics initially depends on preexisting conditions, but the degree may vary across conflicts and time.

Admittedly, the relative location of each class of factors can vary and change in different conflicts. For example, ethnic geography can change through security-motivated migration and forced resettlement of the local population during conflicts (Steele, 2009; Weidmann and Salehyan, 2013; Zhukov, 2015). Nonetheless, it is highly unlikely that these factors change drastically over short time periods. Indeed, previous studies typically treat them as static or determined *ex ante*, rather than dynamic or *ex post*, determinants of civil war violence (e.g, Humphreys and Weinstein, 2006; Toft and Zhukov, 2015; Weinstein, 2005, 2007; Wood, 2014a). Reflecting on these insights, it is reasonable to assume that organizational configurations, external support, ethnic geography, and inter-group competition lie somewhere between the two ends. These factors are likely to be less exogenous to conflict processes than physical geography but less endogenous

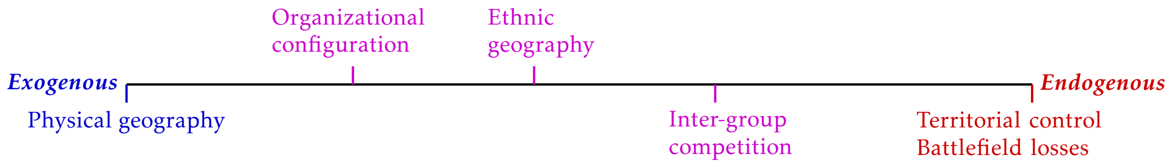


Figure 1: Continuum of determinants of civil war violence

Note: Each class of determinants of civil war violence is ordered according to the extent to which it can be assumed to be exogenous to the conflict process from the left end. The leftmost class of factors includes geographic conditions such as elevation and distance from national capitals, whereas the most dynamic factors, including balance of territorial control and battlefield dynamics, are located at the right end. The exact locations of intermediate factors can vary across conflicts and time.

than the balance of territorial control and battlefield dynamics.

2 Determinants of Violence Depend on the Types of Violence

A noteworthy aspect of existing studies lies in their different formulations of independent and dependent variables. Studies on the determinants of selective violence tend to focus on the effects of endogenous factors, while those on indiscriminate violence highlight the role of exogenous factors. For example, Kalyvas's (2006) theory highlights the impact of territorial control, which is largely endogenous to conflict dynamics, on locations of selective violence. By contrast, Fjelde and Hultman (2014) and Schutte (2017)each demonstrate the vital role of exogenous factors in determining the frequency of indiscriminate violence. Although indiscriminate violence is no more than a product of lag for Kalyvas (2006), these studies suggest that largely static characteristics such as physical geography substantially shape the frequency and locations of this type of violence.

What remains relatively unclear in the literature, however, is the relative importance of exogenous and exogenous factors in shaping the risk of violence in civil conflicts. Building upon the contributions of previous studies, this chapter proposes a nuanced and unified theoretical framework that specifies the likely impacts that the two classes of factors each have on the likelihood of selective and indiscriminate violence.

2.1 Determinants of Indiscriminate Violence

We argue that the relative importance of each class of factors varies depending on the types of violence perpetrated by the warring actors. Specifically, we posit that exogenous factors substantially shape the risk of both selective and indiscriminate violence. We also expect endogenous factors to be less important in determining the locations of indiscriminate violence. Rather, this class of factors plays an important role in altering the locations of selective violence.

Underlying these expectations is the speculation that different types of violence are motivated by different sets of warring parties' incentives. For example, Balcells (2011) and Fjelde and Hultman (2014) demonstrate how a preexisting geographic configuration of suspected supporters of the opponents, which is largely determined by pre-war affiliations, shapes how indiscriminate violence unfolds during civil conflicts.

If indiscriminate violence tends to "backlash" and undermine popular support for the perpetrator (Ellsberg, 1970; Kalyvas, 2006), instances of this type of violence would be either a product of error or reflect incentives that differ from the facilitating popular support of local civilians. As clearly formulated in Azam and Hoeffler (2002) and Fjelde and Hultman (2014), an important strategic consideration that motivates this type of violence is to maximize the damage and costs imposed on the opponents and their collaborators. Collective targeting against the suspected supporters of the opponent is often employed in order to undermine the productive capacity of the opponents' constituency, rather than to expand constituent support for the perpetrator or loot local resources (Azam and Hoeffler, 2002; Downes, 2007; Fjelde and Hultman, 2014; Stanton, 2013; Zhukov, 2015; see also, Acosta, 2016; Downes, 2006, 2008). The use of indiscriminate violence may also be efficient at pressuring the opponent into entering negotiations (Hultman, 2009).

Hultman (2007, 2012) and Wood (2014a) demonstrate how warring parties' incentives to employ indiscriminate tactics vary over time, reflecting battlefield dynamics within single conflicts. These arguments suggest that the *frequency* of indiscriminate use of violence at specific localities may vary over time. Nonetheless, the *locations* that are susceptible to this type of violence, or the locations where warring parties would expect to be able to maximize the damage to the opponent, are largely a function of preexisting or exogenous factors such as physical and ethnic geography. We therefore hypothesize:

Hypothesis 1 (Determinants of indiscriminate violence) *Subnational risks of indiscriminate insurgent violence are determined by exogenous factors.*

2.2 Determinants of Selective Violence

Another key insight from previous studies is that the locations of selective violence reflect the dynamic elements of conflict such as levels of territorial control and recent history of battles. While Kalyvas (2006, 132–140) highlights the role of preexisting geographic factors in determining the initial spatial distribution of territorial control, the theory predicts that the balance of territorial control exercised by warring parties is the primary predictor of selective violence. In contrast to collective targeting, the likely motivation underlying the selective use of violence is to maximize popular support and deter defection, thereby strengthening the perpetrator's territorial control within the targeted regions (Eck, 2014; Herbst, 2000; Kalyvas, 2006). Successful use of selective violence may eventually shift local civilians' support for warring parties and thereby cause subsequent changes in territorial control (Kalyvas, 2006, Chap.7). The changes in territorial control in turn alter the locations that are susceptible to selective violence or where warring parties have incentives and opportunities to employ violence selectively in the subsequent periods.

These dynamics suggest that a recent history of violence as well as preexisting conditions should play an important role in predicting the location of insurgent violence, as they reflect the changing levels of territorial control. Although the underlying causal mechanism remains unspecified, empirical assessments of violence patterns in several civil conflicts have highlighted the role of the records of past violence in shaping future violence (Braithwaite and Johnson, 2012, 2015; Hirose, Imai, and Lyall, 2017; Linke, Witmer, and O'Loughlin, 2012; Zammit-Mangion, Dewar, Kadirkamanathan et al., 2012). We therefore expect insurgents' selective targeting to be a function of not only exogenous factors but also the endogenous dynamics of conflict.

Hypothesis 2 (Determinants of selective violence) *Subnational risks of selective insurgent violence are determined by both exogenous and endogenous factors.*

To evaluate the validity of these hypotheses, we rely upon the precise and micro-level records of violent incidents in the ongoing war in Afghanistan and a computational model. The following section provides a brief overview of the empirical data.

3 Data and Empirical Context

This chapter uses the ongoing irregular warfare in Afghanistan as a case to disentangle the determinants of selective and indiscriminate violence in civil conflicts. Following the former Taliban leader Mullah Muhammad Omar's vow to "retake control of Afghanistan" in 2004 (Gall, 2004), the Taliban remnants had regrouped and launched large-scale insurgency by late 2005 (Johnson, 2013, 10–11). Despite the losses and attrition that the Taliban have suffered and the U.S.-led troop "surge," or a massive increase of coalition troops, the counterinsurgency campaign is not yet completed (Farrell and Giustozzi, 2013; Johnson and DuPee, 2012; Johnson and Mason, 2008).

The following empirical analysis relies on the U.S. military internal database called "Significant Activities" (SIGACTs).[4] The SIGACTs are a collection of short summaries of events in relation to the actors involved, casualties, event type, locations, timing, and other related information that have been recorded by individual troops. The SIGACTs event data cover both violent (e.g., IED explosions) and nonviolent (e.g., information provision from civilians) incidents across the country between January 2004 and December 2009, and have been widely used in the civil war literature (e.g., Donnay and Filimonov, 2014; Schutte, 2017; Weidmann, 2015, 2016; Zammit-Mangion, Dewar, Kadirkamanathan et al., 2012).

Because the activities of ISAF and Afghan national forces are likely to be affected by factors other than the local-level determinants of violence illustrated above, the following analysis employs insurgent violence as the primary dependent variable. Of the $76,910$ entries, $52,196$ comprise reports on violent incidents and the remaining $24,714$ are on nonviolent incidents.[5] We aggregated $45,628$ incidents of insurgent-initiated violence to the settlement level using their geo-coordinates ($N_{\text{stl}} = 37,484$).[6]

[4]As in the previous Chapter, the following empirical analysis employs the "Afghan War Diary" (AWD), a subset of the SIGACTs that has been released by WikiLeaks.org.

[5]Although the SIGACTs database offers a rare opportunity for researchers to explore the microdynamics of civil war, it may suffer from potential bias (Donnay and Filimonov, 2014; Weidmann, 2015, 2016). First, there may be a tendency for military troops to under-report the collateral damage caused by their operations. However, this bias is unlikely to cause a serious problem in the following analysis, since the main focus here is on the distribution of insurgent-initiated attacks. Second, the reporting standards for SIGACTs may vary across units and/or have changed over time, possibly resulting in a significant measurement error. This concern is partly alleviated by focusing on the temporally aggregated spatial distribution of violence.

[6]Individual incidents are tagged with the geographically closest settlements. Geocoded settlement dataset is obtained from the USAID "Afghanistan: Settlements." Available at: https://www.humanitarianresponse.info/operations/afghanistan/dataset/

Table 1: Distribution of insurgent violence across population settlements, by event types

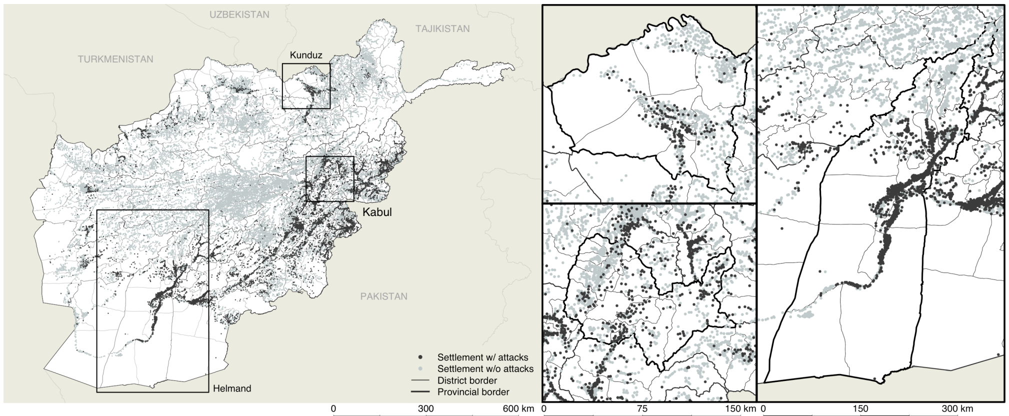
	w/o IED incidents	w/ IED incidents
w/o non-IED incidents	29,840	2,131
w/ non-IED incidents	2,765	2,748

During the period covered by the dataset, 7,644 (20.4%) settlements experienced one or more insurgent attacks.

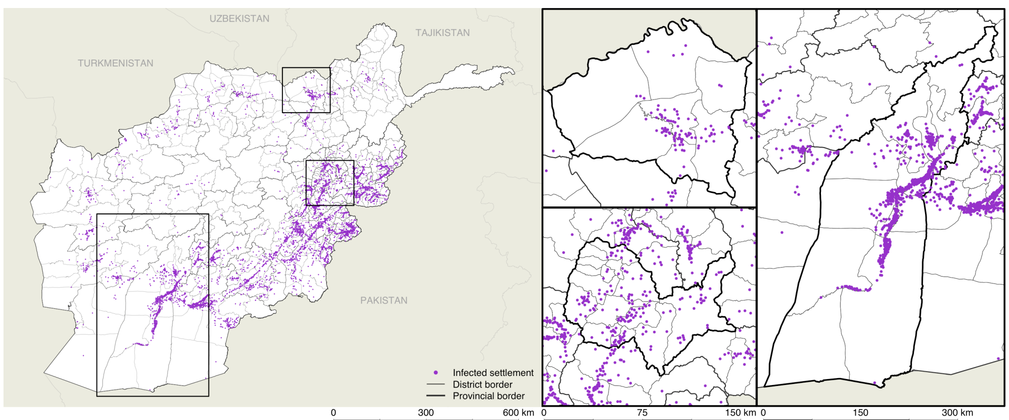
The operationalization of selective and indiscriminate violence is the key to the following empirical analysis. We operationalized indiscriminate insurgent violence by attacks using IED (improvised explosive device) and selective violence by non-IED attacks. Underlying this operationalization is the idea that IED attacks, as exemplified by roadside bombs, are rarely selectively targeted, which generally fits with the operational definition of indiscriminate violence. Specifically, the selection criteria employed the "Affiliation," "Category," and "Type" columns (short event description and perpetrator) in the SIGACTs database to filter the records of IED and non-IED insurgent attacks. Specifically, "Explosive Hazard," "IED Ambush," "IED Explosion," "IED Found/Cleared," "IED Threat," "IED Hoax," "IED False," "IED Suspected," "Interdiction," "Premature Detonation" (premature IED detonation), "Mine Found/Cleared," "Mine Strike," "Unknown Explosion," and "Vehicle Interdiction" categories were coded as IED events, while the remaining events affiliated with insurgents were coded as non-IED events. We further matched the subsets of the data against the "Affiliation" variable, which contains information about the perpetrator ("FRIEND," "ENEMY," "NEUTRAL," "UNKNOWN"), and coded those records with "Affiliation"="ENEMY" as insurgent-initiated events. This coding procedure yielded 19,567 records of indiscriminate IED attacks and 26,061 selective non-IED attacks.

Table 1 summarizes the resultant distribution of IED and non-IED insurgent attacks across villages, and Figure 2 uses maps to visualize the spatial distribution of population settlements with and without insurgent violence. While an initial look at Figure 2 suggests that IED and non-IED attacks tend to cluster in similar regions (panels (b) and (c)), the cross-tabulation reported in Table 1 depicts otherwise. Although 2,748 villages had experienced both types of insurgent attacks, the remaining $2,131+2,765=$

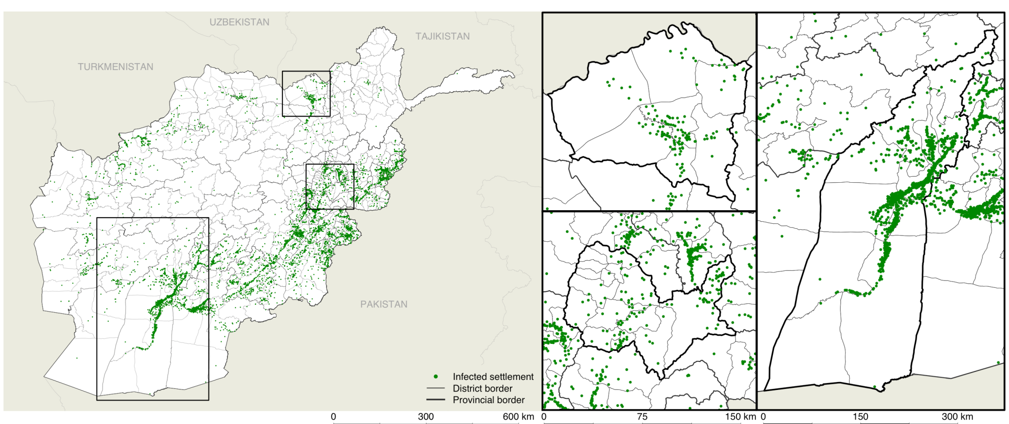
afghanistan-settlements-villages-towns-cities-0, accessed July 25, 2014



(a) Spatial distribution of insurgent violence



(b) Spatial distribution of indiscriminate insurgent violence (IED attacks)



(c) Spatial distribution of selective insurgent violence (non-IED attacks)

Figure 2: Spatial distribution of insurgent violence in Afghanistan, 2004–2009, by event types

Note: (a) Black (●) and gray dots (●) indicate settlements with and without insurgent attacks, respectively. (b) Purple dots (●) represent settlements with indiscriminate (IED) insurgent attacks. (c) Green dots (●) represent the location of settlements with selective (non-IED) insurgent attacks.

4,896 out of 7,644 settlements with one or more insurgent attacks had been exposed to *either* IED or non-IED but not another type of attack during the study period. Similarly, the village-level correlation between the number of IED and non-IED attacks remains modest, with Pearson's $r = 0.418$.[7] The variations in the spatial distributions of insurgent violence offer a suitable foundation for testing the validity of the two propositions advanced in the previous section.

4 Computational Model

We developed a minimal agent-based model to evaluate the plausibility of the two propositions discussed above. Agent-based modeling (ABM) is a computational method that specifies, as a set of computer codes, hypothesized mechanisms that govern the behavior and interactions of constituent elements of a system, commonly called agents. One utility of this computational approach is its ability to represent spatially situated and locally interacting agents, which enables us to model the associations between local conditions and insurgent behavior (de Marchi and Page, 2014). This flexibility is essential in the current context, as our propositions focus on how local-level conditions influence insurgents' incentives and opportunities to engage in violence (Buhaug and Rød, 2006; Zhukov, 2012).

Another utility of agent-based models lies in its flexibility to be incorporated with empirical data, which allows for models to be seeded and validated using observed records. For example, Lim, Metzler, and Bar-Yam (2007) and Weidmann and Salehyan (2013) have developed agent-based models incorporated with spatial data of ethnic geography. Employing the computational method, these studies examine the microfoundations underlying the observed associations between ethnic segregation and violence in India, Iraq, and former Yugoslavia. These models demonstrate that a simple mechanism of ethnically and/or security motivated migration and subsequent violence accounts for the spatial distribution of violence in actual conflicts (see also, Bhavnani, Donnay, Miodownik et al., 2014). Incorporating a two-dimensional model space with the real geography of Afghanistan, the empirically-grounded, agent-based modeling approach allows us to explore how local conditions shape insurgent behavior.

[7] The absence of the lack of spatial overlap between selective and indiscriminate insurgent violence is consistent with the earlier finding of Kalyvas (2006).

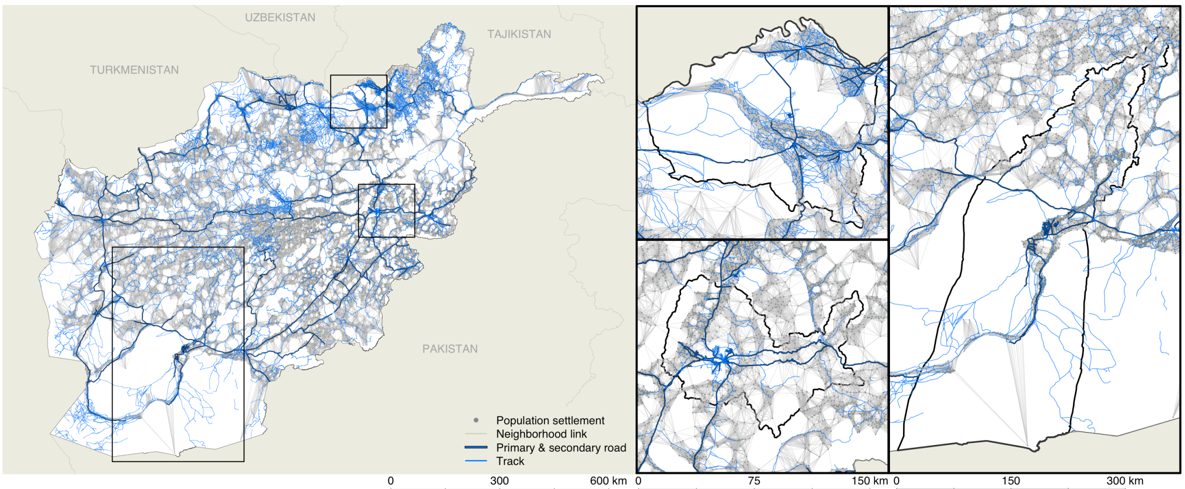


Figure 3: Population settlements and neighborhood network

Note: Gray dots (•) represent individual population settlements, whereas gray segments linking settlements indicate pairs of neighbor settlements. Blue segments indicate road networks.

4.1 Model Space

The model space is specified by a network between population settlements, which mimics the micro-geography of Afghanistan. There is a set $\boldsymbol{S} = \{S_1, \ldots, S_N\}$ of N settlements populated by M insurgent agents $I_j \in \boldsymbol{I}$, with $\boldsymbol{I} = \{I_1, \ldots, I_M\}$ denoting the set of agents. Individual settlements are located according to the corresponding geo-locations and are linked through neighborhood ties, with $\mathcal{S}_i$ denoting the set of neighbor settlements of S_i. Figure 3 illustrates the model space.

Agents are randomly distributed to settlements at the beginning of a simulation run. At every discrete time period $t \in [1, t_{\max}]$, agent I_j makes a binary decision whether to conduct an attack in its current location S_i or relocate to a neighbor settlement $S_l \in \mathcal{S}_i$. Once all agents have made their decisions, the model records the location and number of attacks at t and then proceeds to period $t+1$. The output of a simulation run is a vector of cumulative numbers of insurgent attacks that have been conducted in individual settlements, $\hat{\boldsymbol{Y}} = (\hat{Y}_1, \ldots, \hat{Y}_N)$.

The neighborhood network, or an $N \times N$ spatial weight matrix $\boldsymbol{W}$, represents the pathways through which agents migrate. We define $\boldsymbol{W}$ as a distance-weighted k-nearest neighbor (DWkNN) matrix. The non-diagonal elements $w_{il}^{\text{Spat}} \geq 0$ capture the geographically-weighted influence of settlement S_l on S_i, with diagonal elements

$w_{ii}^{\text{Spat}} = 0$. We first construct a 20-nearest neighborhood matrix ($k = 20$) with $37,484 \times 20 = 749,680$ non-zero entries, in which the 20 geographically closest settlements are defined as neighbors of S_i. We then compute a spatial weight w_{il}^{Spat} for each neighbor pair by measuring the inter-settlement geodesic distances penalized by the additional distances to the nearest roads from individual settlements. For simplicity, we rely on the inverse-distance weighting (IDW) scheme: $w_{il}^{\text{Spat}} = d_{il}^{-\phi}$, where d_{il} indicates the penalized distance between S_i and S_l, and $\phi > 0$ denotes the distance weight.[8]

4.2 Algorithm

The specification of insurgent behavior is a generalization of the model in Weidmann and Salehyan (2013). The original model assumes the probability of attack to be conditional on the local ethnic configuration. The parsimonious model specification provides us with a suitable baseline to model insurgent behavior as a function of local-level conditions.[9] This chapter extends the model such that it incorporates diffusion effects as well as structural factors. Specifically, our model assumes the decision of insurgent I_j located at S_i to carry out an attack at time period $t+1$, or $v_{ijt+1} = 1$ with $v_{ijt} \in \{0, 1\}$, to be a realization of a Bernoulli trial with probability p_{ijt+1}:

$$v_{ijt+1} \sim \text{Bernoulli}\left(p_{ijt+1}\right), \tag{1}$$

$$p_{ijt+1} \equiv \Pr\left(v_{ijt+1} = 1 | \boldsymbol{x}_i, \boldsymbol{z}_{it}\right) = \Lambda\left(\alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta} + \boldsymbol{z}_{it}^\top \boldsymbol{\gamma}\right), \tag{2}$$

where $\Lambda(x) = \exp(x)/(1 + \exp(x))$ is the inverse logit, and α denotes a time- and unit-invariant model parameter. At every time period t, I_j decides to conduct an attack in settlement S_i at $t+1$ with probability p_{ijt+1}; otherwise, I_j relocates to a randomly chosen neighbor settlement $S_l \in \boldsymbol{\mathcal{S}}_i$.

[8] For example, if the inter-settlement geodesic distance between S_i and S_l is 10km and settlement-to-road distances are 1km and 1km, respectively, then $d_{il} = 10+1+1 = 12$km, and $w_{il}^{\text{Spat}} = d_{il}^{-\phi} = 12^{-1} \sim 0.083$ (when $\phi = 1$). The mean (median) value of d_{il} is 9.431 (7.254), and the maximum (minimum) value is 176.3 (0.167) with the standard error of 7.186. Note also that spatial weight w_{il}^{Spat} is rescaled to the range $[0, 1]$. The DWkNN scheme strikes a practical balance between computational feasibility and nuanced approximation of real-world geography. Employing the road network as the neighborhood structure is technically possible. Nonetheless, it is not computationally feasible to construct and run our model on a $37,484 \times 37,484$ origin-destination matrix with $1,405,050,256$ non-zero entries. Although the DWkNN scheme relies on an arbitrary neighborhood size k, it allows for nuanced representation of the neighborhood structure and inter-settlement accessibility at a relatively low computational cost.

[9] See Bhavnani, Donnay, Miodownik et al. (2014) and Siegel (2011) for similar model specifications.

The underlying intuition is that the probability of insurgent violence depends on two classes of factors: first, the inherent and structural *susceptibility* of the settlements where insurgents are located, and second, spatial and temporal *contexts* of conflict. $\boldsymbol{x}_i$ is a vector of settlement-specific, time-invariant covariates (e.g., geographic condition) whereas $\boldsymbol{z}_{it}$ is a vector of time-varying covariates (e.g., local history of violence). The susceptibility parameter $\boldsymbol{\beta}$ and diffusion parameter $\boldsymbol{\gamma}$ are the corresponding coefficient vectors. $\boldsymbol{\gamma}$ can be collapsed into γ_1 and γ_2, which capture the horizontal (spatial) and vertical (temporal) diffusion effects, respectively. Apparently, our model collapses to the original model in Weidmann and Salehyan (2013) when $\boldsymbol{\gamma} = \mathbf{0}$.

4.3 Micro-Mechanisms

The model includes seven $\boldsymbol{x}$ covariates: population size (*PopSize*), Pashtun population size (*PashtunPop*), local income level (*Development*), ruggedness of terrain (*Ruggedness*), distance from roads (*Road*), distance from the capital (*CapDist*), and distance from the Pakistan border (*APborder*).[10] Tables **??** and **??** in Appendix **??** report the summary statistics. Similarly, the model incorporates two $\boldsymbol{z}$ covariates: spatial lag *Spread* and temporal lag *History*. For a given settlement S_i, $Spread_{it}$ is defined as $Spread_{it} = \sum_{j=1}^{k} w_{il}^{\text{Spat}} \hat{y}_{kt}$, where w_{il}^{Spat} is a spatial weight as defined above, and $\hat{y}_{kt}$ denotes the number of insurgent attacks that have occurred at neighbor settlement $S_l \in \mathcal{S}_i$ at time period t. $Spread_{it}$ weights spatially proximate incidents more heavily than remote ones and defines the local spillover effects across immediate neighborhood networks. $History_{it}$ is defined as the temporally weighted sum of insurgent attacks conducted at S_i until period t: $History_{it} = \sum_{\tau=1}^{t} w_{\tau}^{\text{Temp}} \hat{y}_{i\tau}$, where $w_{\tau}^{\text{Temp}} = (t - \tau + 1)^{-\phi}$ is an exponential weight analogous to w_{il}^{Spat}. $History_{it}$ captures the temporally weighted severity of past insurgent activities at S_i, which may also shape the context in which future insurgent activities unfold.

$\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ parameters allow for a nuanced operationalization of the two propositions reviewed in the previous section. $\boldsymbol{\beta}$ parameters govern how local structural factors

[10] *Road*, *CapDist*, and *APborder* are measured by the geodesic distance in kilometers. Data on local income level were derived from the "Geographically based Economic data" (G-Econ, Nordhaus, 2006, http://gecon.yale.edu, accessed July 25, 2014), and other settlement-level attributes are derived from the USAID dataset. *Ruggedness* was computed by taking the elevation variance of the 0.05° resolution grid-cell where the corresponding settlement is located and immediate neighbor cells using SpatialGridBuilder (Pickering, 2016). All $\boldsymbol{x}$ covariates are log-transformed and rescaled to the interval $[-1, 1]$ to minimize the effect of extreme values and make estimates easily comparable.

Table 2: Model parameters

	Symbol	Description/baseline value (robustness checks)
Model setting		
# iteration	$t_{\max}$	300
# population settlements	N	$37,484$
# insurgent agents	M	$20,000$ $(18,000, 22,000)$
# neighbor settlements	k	20 $(10, 30)$
distance decay weight	ϕ	1 $(0.5, 2)$
Insurgent behavior		
Constant	α	
Susceptibility parameter	$\boldsymbol{\beta}$	Coefficient parameters of $\boldsymbol{x}$ covariates
Diffusion parameter	$\boldsymbol{\gamma}$	Coefficient parameters of $\boldsymbol{z}$ covariates

shape insurgent behavior, while $\boldsymbol{\gamma}$ parameters determine how local history of insurgent activities influence future insurgent violence and thus specify the diffusion process. A positive estimate of a given parameter indicates that the corresponding covariate positively (negatively) impacts the settlement-level probability of insurgent violence (relocation). Table 2 summarizes the model parameters.

5 Results I: Determinants of Insurgent Violence

The following two sections report the main findings derived from the computational model. Because the model is not analytically tractable, the analysis derives its results via computational simulation. The validation strategy is twofold: first, we specify empirically plausible parameter sets and thereby examine the likely determinants of insurgent activities. We then evaluate the predictive power of the calibrated model. The analysis in this section aims to optimize the model's parameter combinations such that the simulation outcomes closely fit the empirical records along the specified dimensions of agreement, thereby identifying the likely determinants of insurgent violence. In the following subsections, we first present the validation strategy and then examine the individual parameters estimates.

5.1 Parametrization strategy

Our parametrization strategy broadly follows that of Weidmann and Salehyan (2013, 58). First, $N_{\text{run}} = 50,000$ simulations are conducted with parameter combinations drawn from uniform distributions (parameter space $\mathbf{\Theta}_0$). We then select a subset of parameter combinations $\mathbf{\Theta}_1 \subset \mathbf{\Theta}_0$ that generates spatial distributions of insurgent violence similar to the observed distribution. This parametrization strategy allows for the parameter values to be specified that are necessary to generate realistic patterns of violence and their impacts on simulation outcomes.[11]

Define "good-fit" runs The generated distributions of violence are compared with the empirical records along two target classes: *location* and *number* of insurgent violence. The agreement between the predicted and observed locations of violence is quantified by true positive rate (TPR $= \frac{\#\,\text{true positives}}{\#\,\text{total positives}}$), false positive rate (FPR $= \frac{\#\,\text{false positives}}{\#\,\text{total negatives}}$), and accuracy (ACC $= \frac{\#\,\text{true positives} + \#\,\text{true negatives}}{\#\,\text{total cases}}$). Similarly, the degree of agreement for the number of attacks is quantified by the Root Mean Squared Error (RMSE) $= \sqrt{\sum_{i=1}^{N}(\hat{Y}_i - Y_i)^2/N}$.

We define "good-fit" runs as those that minimize the deviation of the model outcome from the empirical records that fulfill the following conditions: (1) ACC > 0.67, (2) TPR $>$ FPR, and (3) (weighted) RMSE $< \text{RMSE}_{0.05}^{\text{rnd}}$. In order to filter the runs that fulfill these conditions, we first discard the noninformative runs that generated no insurgent attacks and then select those that fulfill these three conditions to obtain the optimized parameter space $\mathbf{\Theta}_1$. A random coin toss produced an ACC score of 0.5, and thus a > 0.67 ACC ensures that the corresponding run correctly classifies more than two-thirds of the observations (condition 1). Similarly, as a general rule, a model with high binary predictive capability has a TPR that is consistently higher than the corresponding FPR (condition 2).

A "good-fit" run should also minimize the deviation of predicted numbers of violence from the observed data series (condition 3). $\text{RMSE}_{0.05}^{\text{rnd}}$ denotes the 5th percentile value of the RMSE distribution obtained by N_{run} random null predictions. A null prediction is generated by assigning the observed number of IED and non-IED attacks to randomly

[11] This parametrization approach allows for a large parameter space to be examined at a relatively low computational cost compared to the oft-employed sequential parameter sweeping that is known to be the equivalent of comparative statics in game-theoretic models (Holland, Holyoak, Nisbett et al., 1989).

selected population settlements. This procedure is repeated N_{run} times to generate a hypothetical sample of "random conflicts." If the RMSE obtained from a run is smaller than $\text{RMSE}^{\text{rnd}}_{0.05}$, the prediction is considered to outperform random guesses.

One concern regarding the reliance on RMSE is that this metric may potentially be ill-suited for the validation here, given that the occurrence of violence is relatively rare in our dataset (13% for IED attacks and 14.7% for non-IED attacks, respectively). A noninformative prediction, which simply assigns $\hat{Y}_i = 0$, would produce a small RMSE indicating a "good-fit." To address this concern, we employ the *Weighted* RMSE (WRMSE) $= \sqrt{\sum_{i=1}^{N} w_i(\hat{Y}_i - Y_i)^2 / \sum_{i=1}^{N} w_i}$, with weight $w_i = 1 - p(Y_i \geq 1) = 0.870$ for the settlements with IED attacks and $w_i = 0.130$ for those without IED attacks (0.853 and 0.147, for non-IED attacks) instead of standard RMSE in the following analysis. As the adjusted Brier score employed in Chadefaux (2014, 15), WRMSE penalizes prediction errors for rare observations (i.e., $Y_i \geq 1$) more severely than those for abundant ones (i.e., $Y_i = 0$).

Detect "significant" parameters The difference between uniform distribution (Θ_0) and the optimized parameter distribution (Θ_1) provides an intuitive indicator of the effects of individual parameters on the model's predictive performance. A significant difference between the parameter values in Θ_0 and Θ_1 indicates a systematic impact of the corresponding parameter on the model's fits, while an insignificant difference indicates otherwise (Weidmann and Salehyan, 2013, 58–60).

A formal statistical test is informative to quantify the resultant difference between the "prior" (Θ_0) and "posterior" (Θ_1) distributions. Nonetheless, because we are interested not only in *whether* there is a statistically significant difference between the two distributions, but also in *how* the distributions differ, standard statistical tests comparing central tendencies, such as the Student's t-test, do not suffice for the purpose here. Indeed, a pair of distributions can significantly differ from each other in their lower or upper tails even when the difference in their central tendencies remains statistically indistinguishable from zero.

To accomplish this task, we employ the Harrell–Davis quantile estimator in conjunction with a percentile bootstrap (Harrell and Davis, 1982; Wilcox, Erceg-Hurn, Clark et al., 2014). This newly proposed estimator quantifies the difference between two given distributions using the differences in paired decile values, and then computes the confidence intervals of the decile differences via a bootstrap estimation while controlling over the Type I (α) error probability. By comparing the paired decile values, this estimator allows us to evaluate whether and in which part (decile) there are statistically significant differences between the two distributions.

Recall that our central theoretical claim expects the determinants of selective and indiscriminate violence to be distinct from each other. If this theoretical expectation is consistent with the empirical records of insurgent violence in Afghanistan, different sets of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ parameters should exhibit significant shifts in optimized parameter space Θ_1 from the population of random distributions.

5.2 Estimation Results

For the following exercise, two sets of N_{run} simulation runs were conducted using parameter combinations $(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$ randomly drawn from uniform distributions $U(-10, 10)$ and different random seeds for two prediction targets (i.e., IED and non-IED attacks). $M = 20,000$ agents are allocated to randomly selected population settlements at the beginning of a run. Each run continues until either (1) t reaches $t_{\max} = 300$, or (2) the cumulative number of simulated insurgent attacks reached the observed number of attacks ($N^{\text{IED}}_{\text{attack}} = 19,567$ for IED attacks and $N^{\text{NonIED}}_{\text{attack}} = 26,061$ for non-IED attacks).[12] Of $50,000$ randomized trials, $4,685$ runs (9.37%, IED attacks) and $1,707$ runs (3.41%, non-IED attacks) fulfilled the criteria above, respectively.

Determinants of IED (indiscriminate) attacks Each panel in Figure 4 represents a layer of information. The first comprises the distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ parameters in $\boldsymbol{\Theta}_1$ that successfully generate realistic spatial patterns of insurgent violence (top density plot). The second comprises the density estimate of the baseline of uniform parameter distribution $\boldsymbol{\Theta}_0$ for comparative purposes (middle density plot). The third comprises quantile difference estimates accompanied by bootstrapped confidence intervals plotted at the bottom of each panel. Using the Harrell–Davis quantile estimator, the third part of each panel quantifies how much the decile values of one distribution (parameter values in $\boldsymbol{\Theta}_0$) need to be arranged to match the other distribution ($\boldsymbol{\Theta}_1$).[13] In other words, the quantile estimator indicates the decile differences between the parameter values in the

[12]The second condition is an arbitrary one to speed up simulation runs. Removing this condition does not markedly alter the results reported below.

[13]The 95% confidence intervals were obtained via 200 bootstraps. WRS package in R (https://github.com/nicebread/WRS) was used to obtain the reported statistics.

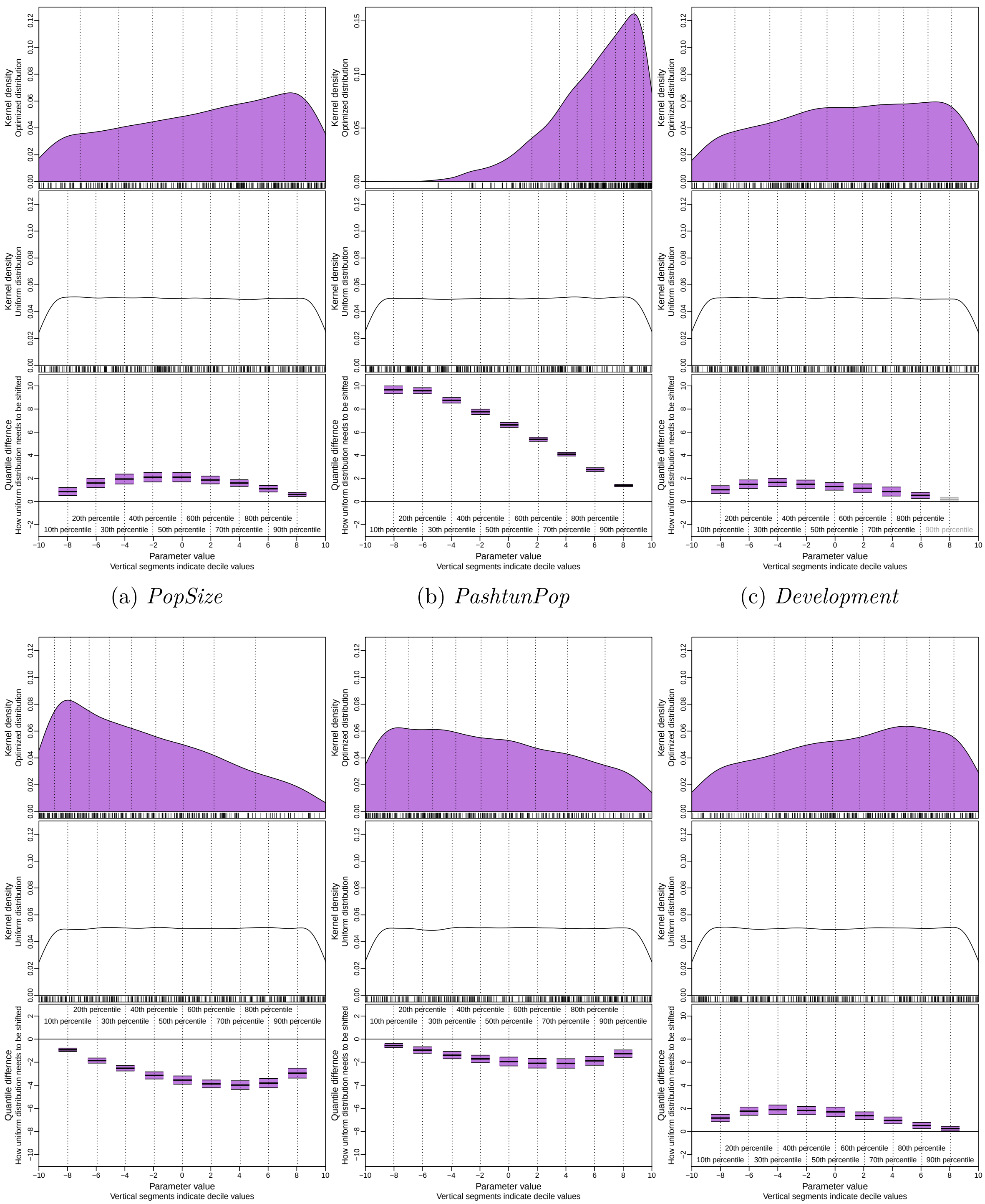
optimized distribution and those in the uniform distribution. A statistically significant shift at the conventional 5% level in each decile is marked by black horizontal segments, while an insignificant shift is shown in light gray.

Hypothesis 1 expects insurgents' decision to employ violence indiscriminately as a function of static factors. The simulation results reported in Figures 4 provide strong support for this theoretical expectation. The effects of this class of predictor of violence are captured by $\boldsymbol{\beta}$ parameters in our computational model. As shown in panels (a) to (g) of Figure 4, almost all $\boldsymbol{\beta}$ parameters are significantly shifted from the population of uniform distribution in the optimized parameter space.

The most apparent impact is found for *PashtunPop* (β_2), which measures the impact of the local Pashtun population on the risk of indiscriminate violence. The 10th to 40th quantiles of the distribution of β_2 in $\boldsymbol{\Theta}_1$ are shifted by more than 8 from uniform distribution, indicating a strong positive impact of *PashtunPop* on the probability of indiscriminate violence. Put another way, this result demonstrates that insurgent agents tend to conduct indiscriminate attacks in settlements with a large Pashtun population in good-fit simulation runs that generate a realistic spatial distribution of indiscriminate violence. Although a simple statistical test comparing the central tendency can also demonstrate a statistically significant difference between the two distributions, it can tell us little about how and how much the distributions differ.

Although similar statistically significant decile shifts are also found for other $\boldsymbol{\beta}$ parameters, the estimated shift sizes remain relatively smaller. Perhaps an exception is *Ruggedness*, or the local elevation differences. The parameter values are negatively skewed in optimized parameter space $\boldsymbol{\Theta}_1$, indicating that insurgent agents are more likely to conduct attacks in easily accessible, rather than inaccessible, settlements (Figure 5(d)). Again, the estimated shift function suggests a statistically significant difference with a relatively large effect size across the parameter range.

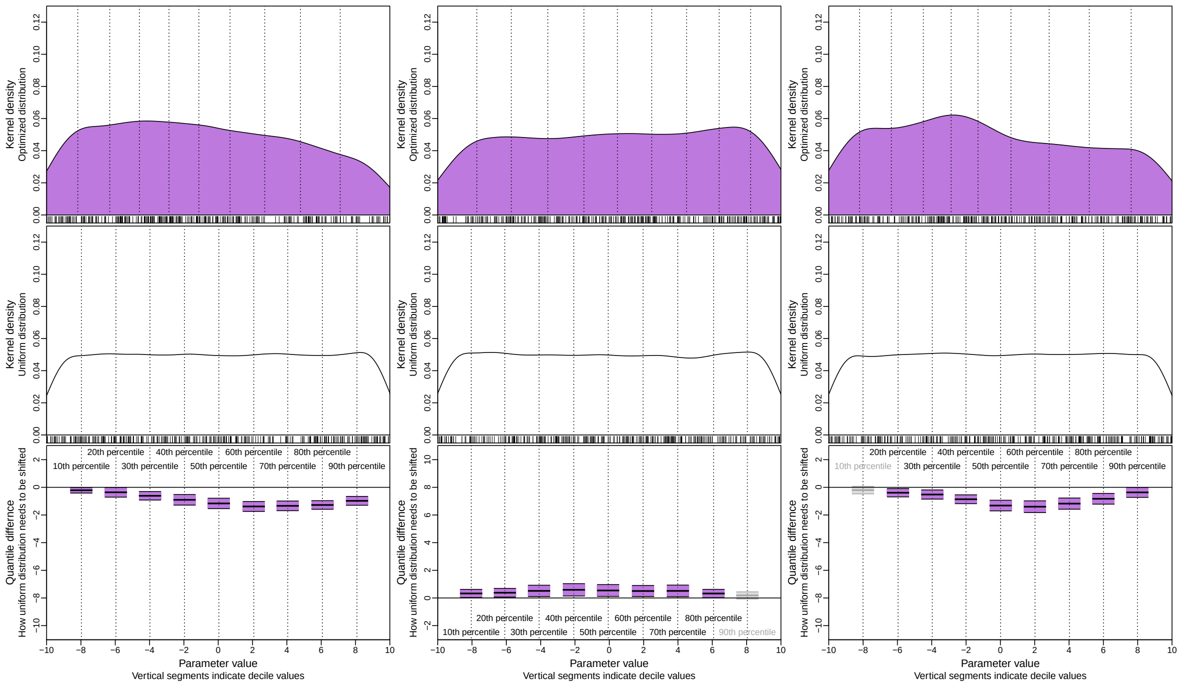
In sharp contrast, the estimated shifts in $\boldsymbol{\gamma}$ parameters, or endogenous factors, remain indeterminate compared with $\boldsymbol{\beta}$ parameters. As illustrated in Figures 4(h) and (i), although many of the estimated shifts of *Spread* and *History* retain statistical significance at the conventional 5% level, the effect sizes remain smaller than the shifts in $\boldsymbol{\beta}$ parameters and substantially insignificant. Combined, these simulation results suggest that insurgent agents' decisions to employ indiscriminate IED attacks are largely a function of exogenous factors, while endogenous factors or a recent history of violence have little impact on the risk of this type of insurgent attack.



(a) *PopSize* (b) *PashtunPop* (c) *Development*

(d) *Ruggedness* (e) *RoadDist* (f) *CapDist*

Figure 4: Optimized parameter distribution, IED attacks



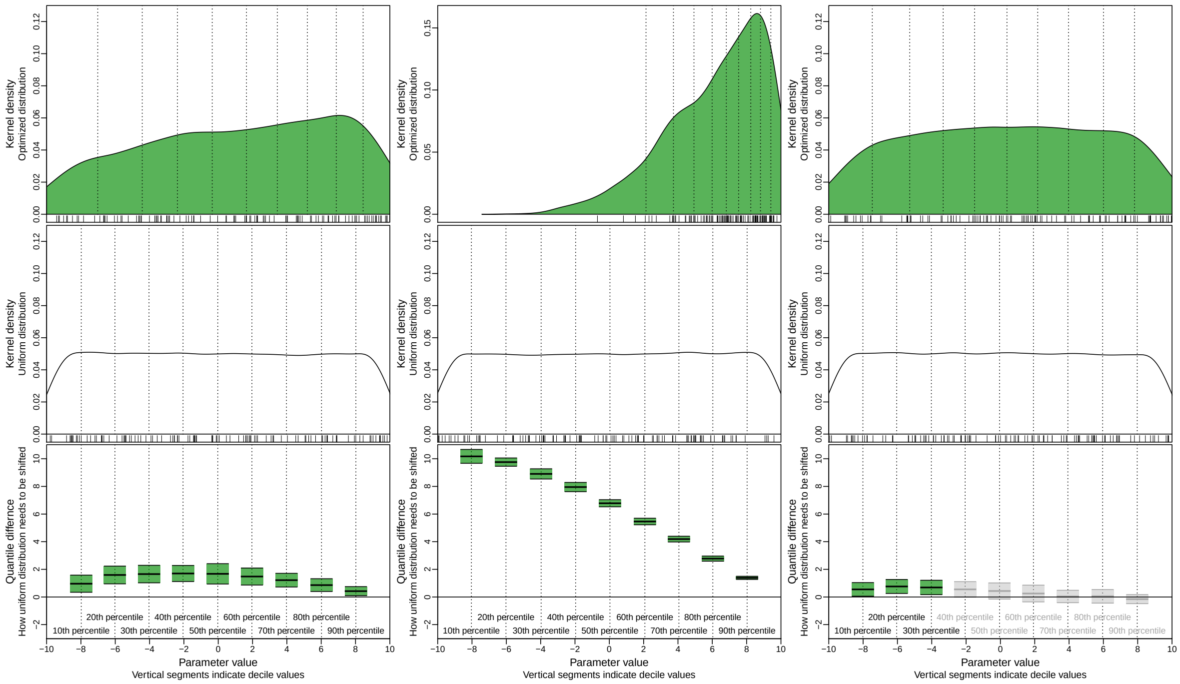
(g) (cont.) *APborder* (h) (cont.) *Spread* (i) (cont.) *History*

Figure 4 (cont.): Optimized parameter distribution, IED attacks

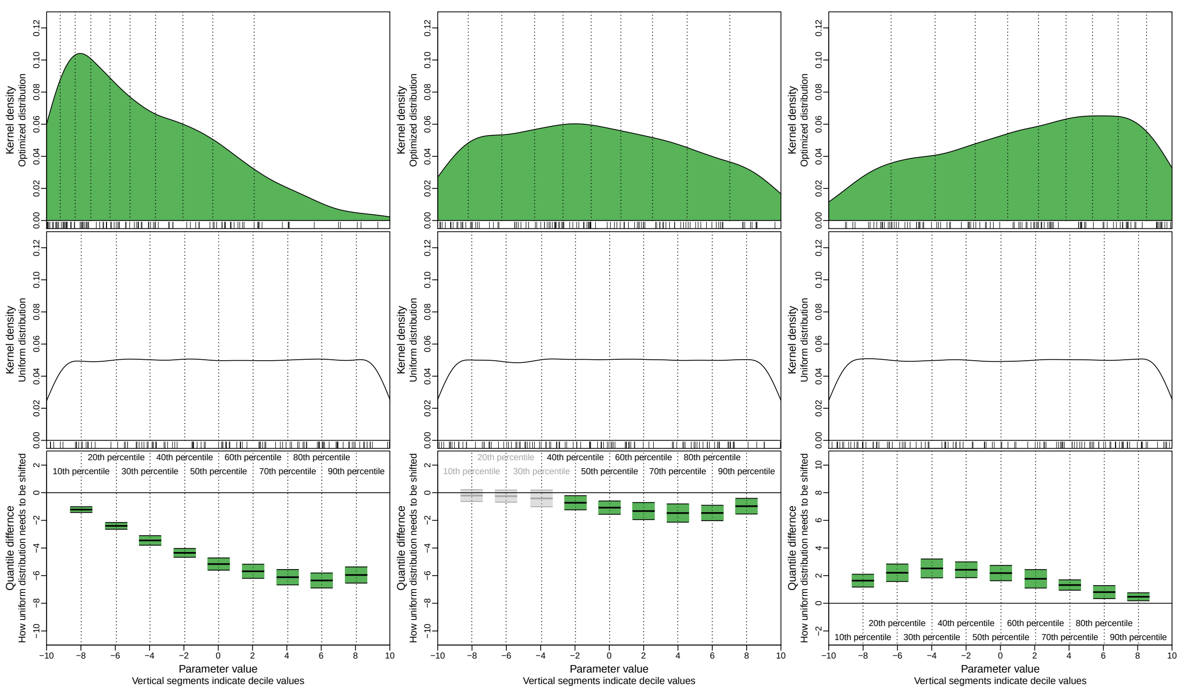
Note: The topmost raw in each panel represents the density estimate for a given parameter in Θ_1, while the middle raw shows the corresponding density in Θ_0. Vertical dashed segments indicate the decile values of each parameter in Θ_0 and Θ_1. The bottom raw plots the decile shift estimates. The decile-difference estimates (thick horizontal segments) between the optimized and uniform distributions are plotted along the vertical axis for each decile of uniform distribution. Thin horizontal segments and gray shades indicate the corresponding 95% bootstrap confidence intervals. Significant differences at the 5% level are marked by black segments, while insignificant differences are marked by gray segments.

Determinants of non-IED (selective) attacks Figure 5 represents the decile-shift plots for the determinants of non-IED attacks. Two significant results emerge from Figure 5. First, the two exogenous factors, *PashutunPop* and *Ruggedness*, that are found to be strong predictors of IED attacks exhibit clear decile shifts in Figure 5. As the shift signs indicate, *PashutunPop* positively impact the risks of indiscriminate (IED) and selective (non-IED) insurgent attacks while *Ruggedness* negatively impacts the likelihood of both types of insurgent violence.

Second and more importantly, one of the modeled endogenous factors, namely *History*, is found to have a substantial negative impact on the agents' decision to conduct non-IED attacks. The large and consistently negative shifts of γ_2 suggest that a marked

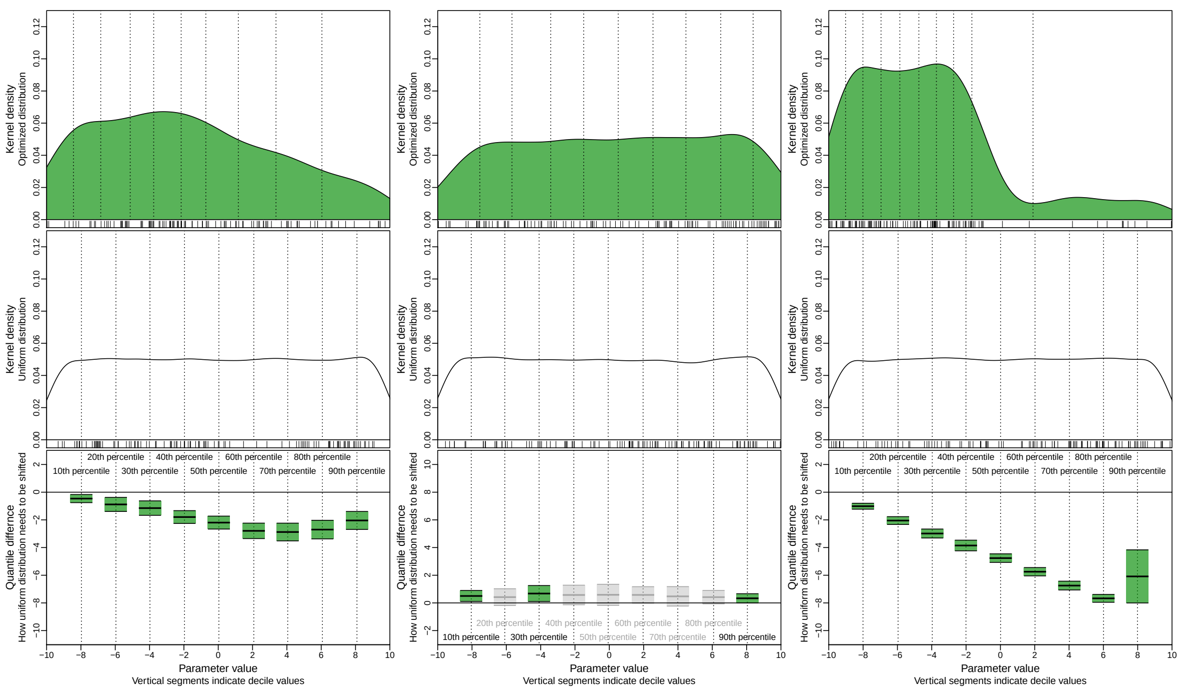


(a) *PopSize* (b) *PashtunPop* (c) *Development*



(d) *Ruggedness* (e) *RoadDist* (f) *CapDist*

Figure 5: Optimized parameter distribution, non-IED attacks



(g) (cont.) *APborder* (h) (cont.) *Spread* (i) (cont.) *History*

Figure 5 (cont.): Optimized parameter distribution, non-IED attacks
Note: See notes in Figure 4.

history of violence facilitates agents' migration to nearby settlements rather than further violence in the originating settlements. The corresponding quantile estimates further indicate the sizable differences in deciles between parameter values in $\mathbf{\Theta}_0$ and $\mathbf{\Theta}_1$. In contrast, the estimate for γ_1 (*Spread*) remains weaker or statistically indistinguishable from the uniform distribution across the sampling range, suggesting that γ_1 is unlikely to have a systematic impact on the model's fit with the empirical records. In other words, the results suggest that while a history of violent activities systematically shapes the future prospects of the violence in a given settlement, spatial context may not matter in determining the probability of insurgent violence.

Generally, the simulation results provide strong empirical support for our central theoretical claim that the determinants of violence vary across types of violence. Recall that Hypothesis 2 posits that insurgents' decisions to employ selective violence is a function of endogenous as well as exogenous factors. Combined with the simulation results for $\boldsymbol{\beta}$ parameters, these results in Figure 5 are consistent with this theoretical expectation.

6 Results II: Prediction Performance

Does the model correctly predict the location and number of insurgent attacks across settlements, and to what extent? What are the determinants of the model's predictive performance? The analysis in the previous section provided valuable insights into the determinants of insurgent violence, yet on its own it provides little information on the veracity of the model. Indeed, the validity of the simulation experiment relies on a potentially unwarranted assumption that the model's explanatory power is at least reasonable. An assessment of predictive performance should be a valuable heuristic in this context (Ward, Greenhill, and Bakke, 2010).

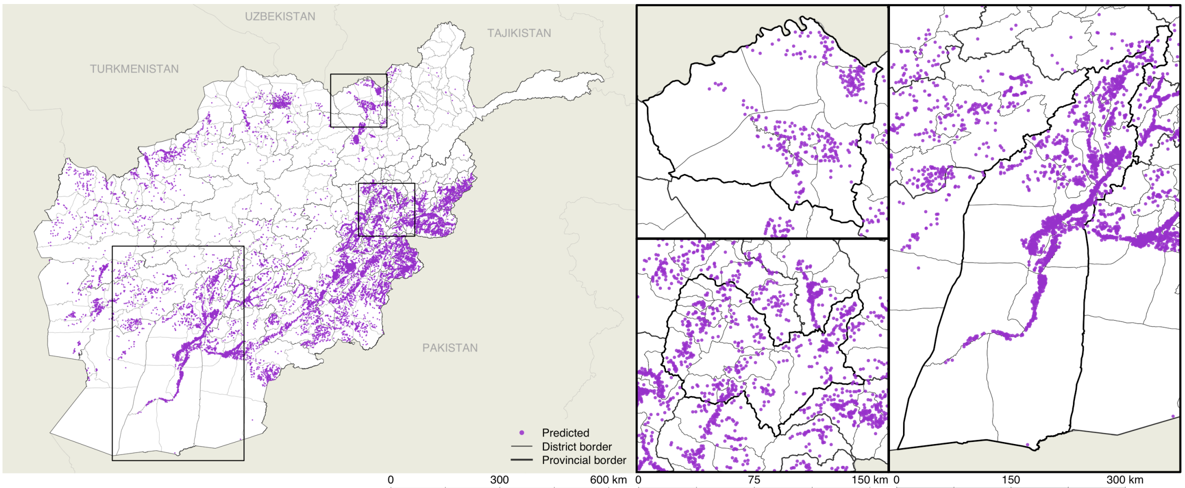
A model's capability to correctly classify binary outcomes (e.g., presence or absence of insurgent violence) can be quantified using the Receiver Operating Characteristic (ROC) curve and the area under the ROC curve (AUC) score. An ROC curve plots TPR and FPR as the output of each possible probability threshold for positive prediction. The resultant plot displays the balance between TPR and FPR, where a highly predictive model (with high TPR and low FPR) produces the curve up in the top left corner. An AUC score, which is defined as the area covered by the corresponding ROC curve, ranges between 0 and 1, and provides a single number summary of the model's classification performance. A random coin toss produces an AUC score of 0.5, whereas a model with higher classification performance should yield an AUC score of greater than 0.5.

Figure 6 maps the (a) predicted locations of IED and (b) non-IED insurgent attacks to visualize the model's predictive performance. The ROC analysis yields AUC scores of 0.794 (95% CI: 0.793, 0.806, IED attacks) and 0.789 (95% CI: 0.785, 0.797, non-IED attacks), indicating that the model's classification performance far exceeds randomness.[14]

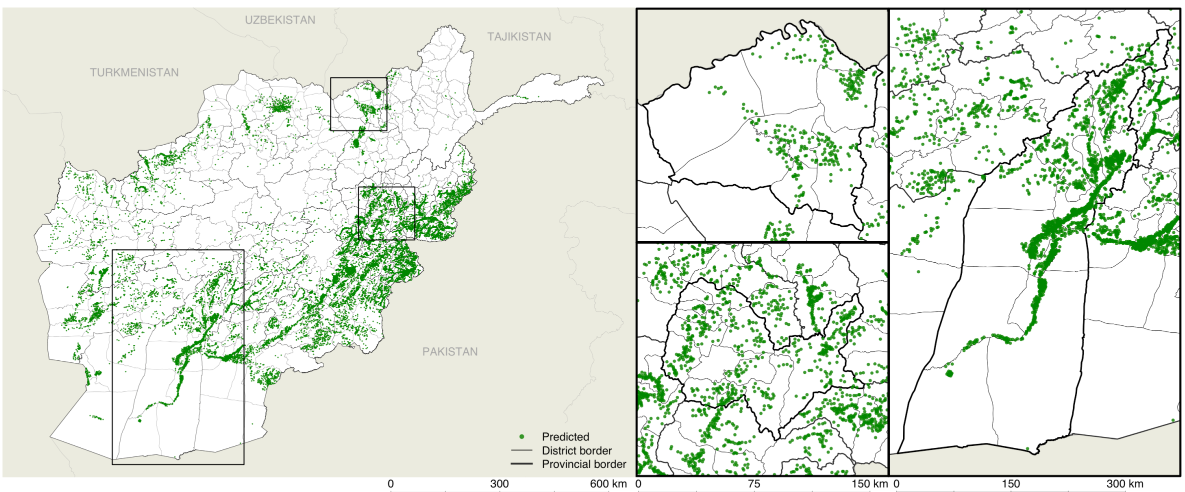
7 Robustness Checks

The main results do not on their own preclude the potential sensitivities of the simulation experiments. Consequently, one may reasonably wonder how the "moving parts"

[14] The predicted probability of violence assigned to each settlement reflects the fraction of simulation runs in $\mathbf{\Theta}_1$ in which one or more attacks have occurred in the corresponding settlement. The 95% CIs were obtained by bootstrap using R's pROC package (Robin, Turck, Hainard et al., 2011).



(a) Predicted locations of IED attacks



(b) Predicted locations of non-IED attacks

Figure 6: Predictive performance: Locations of IED and non-IED attacks

Note: (a) spatial distribution of predicted locations (settlements) of IED attacks. (b) predicted locations of non-IED attacks. These figures are generated using the best threshold values obtained by the ROC analysis.

or parameter settings of the computational model change the results. Four parameters and assumptions warrant investigation to examine the robustness of the main results: (1) neighborhood size k, (2) number of agents M, (3) the attack-or-relocate dichotomy in the behavior algorithm, and (4) exponential weight ϕ for *Spread* and *History*.

To examine the robustness of the main results, $700,000$ additional simulations were conducted, varying these parameter settings and assumption. Reassuringly, none of these sensitivity tests reported in Appendix **??** yielded results that deviate markedly from the main results reported above. These results provide confidence that the specific parameter settings and assumption are not driving the main findings.

8 Conclusion

Civil war studies have increasingly explored the determinants of violence during civil conflicts. Theoretically, the distinction between selective and indiscriminate violence lies at the center of the debate. Empirically, previous studies have demonstrated how a variety of factors can alter the frequency, locations, and types of violence in civil conflicts. Building upon these insights, this chapter has proposed that the determinants of civil-war violence vary across types of perpetrated violence: the decision by warring parties to employ indiscriminate violence is largely a function of exogenous factors, whereas selective violence is a function of endogenous as well as exogenous factors. Drawing on the SIGACTs event data and spatial data of local geography in Afghanistan, the results from the empirically-grounded, agent-based model have yielded two main findings. First, exogenous factors substantially shape agents' decisions to attack indiscriminately. Second, endogenous factors, or a recent history of violence within the same localities, have a sizable impact on agents' decisions to employ selective violence. These results of empirically-based, agent-based simulations provide compelling support for our theoretical argument.

This chapter has significant implications for scholarly and policy debates. First, these findings underscore the importance of disaggregating the types of violence used in civil conflicts. The simulation results demonstrate that while several exogenous or static factors have substantial impacts on the risk of indiscriminate violence, the relative importance of endogenous factors may vary across types of violence. Although no single case study can provide a definitive answer, closer attention, both theoretically and empirically, should be paid to this difference in future research.

Second and methodologically, this chapter demonstrates the methodological utility of data-driven computational modeling. While it is often difficult to disentangle the endogenous and exogenous explanations of the conflict process using observational data alone, the computational approach allows researchers to tackle this challenging task.

Indeed, this chapter demonstrates how the empirically-based computational approach helps us to isolate the impact of each factor and supplements the standard observational approaches.

Finally, this chapter should also inform policymakers and practitioners of counterinsurgency. Counterinsurgency campaigns primarily aim at minimizing insurgent activities and restoring the state's monopoly on violence within its borders. If the determinants of insurgent violence vary across types of violence, effective counterinsurgency campaigns should also vary across targeted types of violence. Thus, rather than adopting a blanket approach, counterinsurgency efforts to contain different types of insurgent violence also need to address different factors if they are to be successful.

References

Acosta, Benjamin. 2016. Dying for survival: Why militant organizations continue to conduct suicide attacks. *Journal of Peace Research* 53 (2):180–196.

Azam, Jean-Paul. 2006. On thugs and heroes: Why warlords victimize their own civilians. *Economics of Governance* 7 (1):53–73.

Azam, Jean-Paul, and Anke Hoeffler. 2002. Violence Against Civilians in Civil Wars: Looting or Terror? *Journal of Peace Research* 39 (4):461–85.

Balcells, Laia. 2011. Continuation of Politics by Two Means: Direct and Indirect Violence in Civil War. *Journal of Conflict Resolution* 55 (3):397–422.

Beardsley, Kyle, and Brian McQuinn. 2009. Rebel Groups as Predatory Organizations: The Political Effects of the 2004 Tsunami in Indonesia and Sri Lanka. *Journal of Conflict Resolution* 53 (4):624–45.

Bhavnani, Ravi, Karsten Donnay, Dan Miodownik, Maayan Mor, and Dirk Helbing. 2014. Group segregation and urban violence. *American Journal of Political Science* 58 (1):226–245.

Boulding, Kenneth E. 1962. *Conflict and defense: A general theory.* New York: Harper Torchbooks.

Braithwaite, Alex, and Shane D. Johnson. 2012. Space-Time Modeling of Insurgency and Counterinsurgency in Iraq. *Journal of Quantitative Criminology* 28 (1):31–48.

———. 2015. The battle for Baghdad: Testing hypotheses about insurgency from risk heterogeneity, repeat victimization, and denial policing approaches. *Terrorism and Political Violence* 27 (1):112–132.

Bueno de Mesquita, Ethan. 2013. Rebel tactics. *Journal of Politcal Economy* 121 (2):323–357.

Buhaug, Halvard, and Jan Ketil Rød. 2006. Local determinants of African civil wars, 1970–2001. *Political Geography* 25 (3):315–335.

Chadefaux, Thomas. 2014. Early Warning Signals for War in the News. *Journal of Peace Research* 51 (1):5–18.

De la Calle, Luis. 2017. Compliance vs. constraints: A theory of rebel targeting in civil war. *Journal of Peace Research* 54 (3):427–441.

de Marchi, Scott, and Scott E. Page. 2014. Agent-Based Models. *Annual Review of Political Science* 17 (1):5.1–5.20.

Dell, Melissa, and Pablo Querubin. 2018. Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies. *Quarterly Journal of Economics* 133 (2):701–764.

Di Salvatore, Jessica. 2016. Inherently vulnerable? Ethnic geography and the intensity of violence in Bosnian civil war. *Political Geography* 51 (1):1–14.

Donnay, Karsten, and Vladimir Filimonov. 2014. Views to a war: Systematic differences in media and military reporting of the war in Iraq. *EPJ Data Science* 3 (1):25.

Downes, Alexander B. 2006. Desperate Times, Desperate Measures: The Causes of Civilian Victimization in War. *International Security* 30 (4):152–195.

———. 2007. Draining the Sea by Filling the Graves: Investigating the Effectiveness of Indiscriminate Violence as a Counterinsurgency Strategy. *Civil Wars* 9 (4):420–444.

———. 2008. *Targeting Civilians in War*. Ithaca, NY: Cornell University Press.

Eck, Kristine. 2014. Coercion in rebel recruitment. *Security Studies* 23 (2):364–398.

Ellsberg, Daniel. 1970. Revolutionary judo. Working notes on vietnam, no. 10, RAND Corporation.

Farrell, Theo, and Antonio Giustozzi. 2013. The Taliban at war: Inside the Helmand insurgency, 2004–2012. *International Affairs* 89 (2013):845–871.

Fjelde, Hanne, and Lisa Hultman. 2014. Weakening the Enemy: A Disaggregated Study of Violence against Civilians in Africa. *Journal of Conflict Resolution* 58 (7):1230–1257.

Gall, Carlotta. 2004. Taliban Leader Vows Return. *The New York Times*, November 13, 2004.

Harrell, Frank E, and C.E Davis. 1982. A New Distribution-Free Quantile Estimator. *Biometrika* 69 (3):635–640.

Hechter, Michael. 1987. *Principles of Group Solidarity*. Barkley: University of California Press.

Herbst, Jeffrey. 2000. Economic incentives, natural resources and conflict in Africa. *Journal of African Economics* 9 (3):270–294.

Hirose, Kentaro, Kosuke Imai, and Jason Lyall. 2017. Can civilian attitudes predict insurgent violence? Ideology and insurgent tactical choice in civil war. *Journal of Peace Research* 54 (1):47–63.

Holland, John H, Keith J Holyoak, Richard E Nisbett, and Paul R Thagard. 1989. *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Hultman, Lisa. 2007. Battle losses and rebel violence: Raising the costs for fighting. *Terrorism and Political Violence* 19 (2):205–222.

———. 2009. The power to hurt in civil war: The strategic aim of RENAMO violence. *Journal of Southern African Studies* 35 (4):821–834.

———. 2012. COIN and civilian collaterals: Patterns of violence in Afghanistan, 2004–2009. *Small Wars & Insurgencies* 23 (2):245–263.

Humphreys, Macartan, and Jeremy M. Weinstein. 2006. Handling and manhandling civilians in civil war. *American Political Science Review* 100 (3):429–47.

Johnson, Thomas H. 2013. Taliban adaptations and innovations. *Small Wars & Insurgencies* 24 (1):3–27.

Johnson, Thomas H., and Matthew C. DuPee. 2012. Analysing the new Taliban Code of Conduct (Layeha): An assessment of changing perspectives and strategies of the Afghan Taliban. *Central Asian Survey* 31 (December 2014):77–91.

Johnson, Thomas H., and M Chris Mason. 2008. No sign until the burst of fire: Understanding the Pakistan-Afghanistan frontier. *International Security* 32 (4):41–77.

Johnston, Patrick. 2008. The geography of insurgent organization and its consequences for civil wars: Evidence from Liberia and Sierra Leone. *Security Studies* 17 (1):107–137.

Kalyvas, Stathis N. 1999. Wanton and Senseless? The Logic of Massacres in Algeria. *Rationality and Society* 11 (3):243–85.

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

Kalyvas, Stathis N., and Matthew Adam Kocher. 2007. How 'Free' is Free Riding in Civil Wars? Violence, Insurgency, and the Collective Action Problem. *World Politics* 59 (2):177–216.

———. 2009. The Dynamics of Violence in Vietnam: An Analysis of the Hamlet Evaluation System (HES). *Journal of Peace Research* 46 (3):335–355.

Kocher, Matthew Adam, Thomas B. Pepinsky, and Stathis N. Kalyvas. 2011. Aerial Bombing and Counterinsurgency in the Vietnam War. *American Journal of Political Science* 55 (2):201–218.

Leites, Nathan, and Charles Jr. Wolf. 1970. *Rebellion and Authority: An Analytic Essay on Insurgent Conflicts*. Chicago: Markham Publishing Company.

Lichbach, Mark Irving. 1995. *The Rebel's Dilemma*. Ann Arbor: University of Michigan Press.

Lim, May, Richard Metzler, and Yaneer Bar-Yam. 2007. Global pattern formation and ethnic/cultural violence. *Science* 317 (5844):1540–1544.

Linke, Andrew M., Frank D. W. Witmer, and John O'Loughlin. 2012. Space-time Granger analysis of the war in Iraq: A study of coalition and insurgent action-reaction. *International Interactions* 38 (4):402–425.

Lyall, Jason. 2009. Does indiscriminate violence incite insurgent attacks? Evidence from Chechnya. *Journal of Conflict Resolution* 53 (3):331–362.

Merom, Gil. 2003. *How Democracies Lose Small Wars: State, Society, and the Failures of France in Algeria, Israel in Lebanon, and the United States in Vietnam*. Cambridge: Cambridge University Press.

Metelits, Claire M. 2010. *Inside Insurgency: Violence, Civilians, and Revolutionary Group Behavior*. New York: New York University Press.

Nordhaus, William D. 2006. Geography and Macroeconomics: New Data and New Findings. *Proceedings of the National Academy of Sciences of the United States of America* 103 (10):3510–3517.

Olson, Mancur. 1965. *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.

Ottmann, Martin. 2017. Rebel constituencies and rebel violence against civilians in civil conflicts. *Conflict Management and Peace Science* 34 (1):27–51.

Pickering, Steve. 2016. Introducing SpatialGridBuilder: A new system for creating geo-coded datasets. *Conflict Management and Peace Science* 33 (4):423–447.

Raleigh, Clionadh. 2012. Violence Against Civilians: A Disaggregated Analysis. *International Interactions* 38 (4):462–481.

Robin, Xavier, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez, and Markus Müller. 2011. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12 (1):77–84.

Salehyan, Idean, David Siroky, and Reed M Wood. 2014. External Rebel Sponsorship and Civilian Abuse: A Principal-Agent Analysis of Wartime Atrocities. *International Organization* 68 (68):633–661.

Schelling, Thomas C. 1966. *Arms and influence*. New Haven: Yale University Press.

Schutte, Sebastian. 2015. Geography, Outcome, and Casualties: A Unified Model of Insurgency. *Journal of Conflict Resolution* 59 (6):1101–1128.

———. 2017. Geographic determinants of indiscriminate violence in civil wars. *Conflict Management and Peace Science* 34 (4):380–405.

Siegel, David A. 2011. When does repression work? Collective action in social networks. *Journal of Politics* 73 (4):993–1010.

Souleimanov, Emil Aslan, and David S. Siroky. 2016. Random or retributive? Indiscriminate violence in the Chechen Wars. *World Politics* 68 (4):677–712.

Stanton, Jessica A. 2013. Terrorism in the Context of Civil War. *Journal of Politics* 75 (4):1009–1022.

Steele, Abbey. 2009. Seeking Safety: Avoiding Displacement and Choosing Destinations in Civil Wars. *Journal of Peace Research* 46 (3):419–29.

Toft, Monica Duffy, and Yuri M. Zhukov. 2015. Islamists and nationalists: Rebel motivation and counterinsurgency in Russia's North Caucasus. *American Political Science Review* 109 (2):222–238.

Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke. 2010. The perils of policy by p-value: Predicting civil conflicts. *Journal of Peace Research* 47 (4):363–375.

Weidmann, Nils B. 2015. On the accuracy of media-based conflict event data. *Journal of Conflict Resolution* 59 (6):1129–1149.

———. 2016. A closer look at reporting bias in conflict event data. *American Journal of Political Science* 60 (1):206–218.

Weidmann, Nils B., and Idean Salehyan. 2013. Violence and ethnic segregation: A computational model applied to Baghdad. *International Studies Quarterly* 57 (1):52–64.

Weinstein, Jeremy M. 2005. Resources and the Information Problem in Rebel Recruitment. *Journal of Conflict Resolution* 49 (4):598–624.

———. 2007. *Inside rebellion: The politics of insurgent violence*. Cambridge: Cambridge University Press.

Wilcox, Rand R., David M. Erceg-Hurn, Florence Clark, and Michael Carlson. 2014. Comparing two independent groups via the lower and upper quantiles. *Journal of Statistical Computation and Simulation* 84 (7):1543–1551.

Wood, Reed M. 2010. Rebel capability and strategic violence against civilians. *Journal of Peace Research* 47 (5):601–614.

———. 2014a. From Loss to Looting? Battlefield Costs and Rebel Incentives for Violence. *International Organization* 68 (4):979–999.

———. 2014b. Opportunities to kill or incentives for restraint? Rebel capabilities, the origins of support, and civilian victimization in civil war. *Conflict Management and Peace Science* 31 (5):461–480.

Wood, Reed M., and Jacob D. Kathman. 2015. Competing for the Crown: Inter-rebel Competition and Civilian Targeting in Civil War. *Political Research Quarterly* 68 (1):167–179.

Zammit-Mangion, Andrew, Michael Dewar, Visakan Kadirkamanathan, and Guido Sanguinetti. 2012. Point process modelling of the Afghan War Diary. *Proceedings of the National Academy of Science* 109 (31):12,414–12,419.

Zhukov, Yuri M. 2012. Roads and the diffusion of insurgent violence. *Political Geography* 31 (3):144–156.

———. 2015. Population resettlement in war: Theory and evidence from Soviet archives. *Journal of Conflict Resolution* 59 (7):1155–1185.

———. 2017. External resources and indiscriminate violence: Evidence from German-occupied Belarus. *World Politics* 69 (1):54–97.